

NUSC Technical Report 6836  
15 May 1989

AD-A218 133

# A Bayesian Approach to Acoustic Imaging and Object Classification by High Frequency Sonar

J. G. Kelly

R. N. Carpenter

Weapon Systems Technology and Assessment Department

DTIC  
ELECTE  
FEB 22 1990  
S B D  
Co



**Naval Underwater Systems Center**  
Newport, Rhode Island / New London, Connecticut

Approved for public release; distribution is unlimited.

90 02

## PREFACE

This research was conducted under NUSC Project No. A64930, "Arctic Detection and Classification," principal investigator R.N. Carpenter (Code 8212). The sponsoring activity was NUSC's in-house Independent Research and Independent Exploratory Development Program, program manager K.M. Lima (Code 102), and the Office of Naval Research.

The technical reviewer for this report was C.J. Albanese (Code 8212).

The authors wish to thank D.R. Childs (Code 8212) for generating the cumulative distribution functions used to assess classifier performance.

REVIEWED AND APPROVED: 15 MAY 1989



F.L. White  
Head, Weapon Systems Technology  
and Assessment Department

# REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S) TR 6836			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION Naval Underwater Systems Ctr		6b. OFFICE SYMBOL (If applicable) Code 8212		7a. NAME OF MONITORING ORGANIZATION	
6c. ADDRESS (City, State, and ZIP Code) Newport Laboratory Newport, R.I. 02841			7b. ADDRESS (City, State, and ZIP Code)		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of the Chief of Naval Research		8b. OFFICE SYMBOL (If applicable)		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c. ADDRESS (City, State, and ZIP Code) Chief of Naval Research Arlington, VA 22217-5000			10. SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.
			WORK UNIT ACCESSION NO.		
11. TITLE (Include Security Classification) A BAYESIAN APPROACH TO ACOUSTIC IMAGING AND OBJECT CLASSIFICATION BY HIGH FREQUENCY SONAR					
12. PERSONAL AUTHOR(S) Kelly, J.G., and Carpenter, R.N.					
13a. TYPE OF REPORT		13b. TIME COVERED FROM TO		14. DATE OF REPORT (Year, Month, Day) 89-05-15	
				15. PAGE COUNT 95	
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP			
			Active Sonar Systems		
			Target Classification		
			Bayesian Decision Rule		
19. ABSTRACT (Continue on reverse if necessary and identify by block number) The active sonar classification problem is approached as a likelihood ratio test of multiple, alternative hypotheses versus a noise-only null hypothesis. The data are, in general, vector-valued stochastic processes representing measurements from individual elements within a sonar array. An explicit form is assumed for the received signal model, which is statistically characterized for each alternative hypothesis (target class). Explicit results are derived for the likelihood ratio, and various performance characteristics are shown. Moreover, the optimal processor is examined from the perspective of acoustic image processing. Generalizations of the results are indicated and in some cases addressed in detail (e.g., the case of moving targets).					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED		
22a. NAME OF RESPONSIBLE INDIVIDUAL R. N. Carpenter			22b. TELEPHONE (Include Area Code) (401) 841-2406		22c. OFFICE SYMBOL Code 8212

## TABLE OF CONTENTS

	Page
INTRODUCTION .....	1
CLASSIFICATION PROBLEM .....	3
SIGNAL MODEL .....	7
NOISE MODEL .....	11
ESTIMATION OF $\mathbf{a}_i$ .....	13
OPTIMAL BAYES CLASSIFIER .....	16
LIKELIHOOD RATIO .....	19
Representation of $\mathbf{x}(t)$ .....	19
Gaussian Case.....	24
VARIANCE OF $\hat{\mathbf{a}}_i$ AND RESOLUTION.....	31
CLASSIFIER PERFORMANCE .....	46
EXTENSIONS .....	52
Moving Objects.....	52
Non-Gaussian Signal .....	56
CONCLUSIONS.....	58
APPENDIX A: A CONTINUOUS MODEL FOR THE SIGNAL .....	A-1
APPENDIX B: PERFORMANCE FOR THE BINARY CLASSIFICATION PROBLEM .....	B-1
REFERENCES .....	R-1

## LIST OF ILLUSTRATIONS

Figure		Page
1	Signal Model.....	9
2	Nomenclature for Nonrefractive Medium .....	34
3	Example: Two-dimensional Case and Linear Array .....	35
4	Test Regions for Example: Two-dimensional Linear Array .....	37
5	Variance Ratio $\text{tr}[\text{var}(\hat{\mathbf{a}})]/\text{tr}[\mathbf{K}_a]$ vs Angle Cell Width $\Delta\theta$ and Received Signal-to-Noise Ratio $\rho_N \sigma_a^2$ (Two and Eight Angle Cells).....	38



odes  
or

A-1

# LIST OF ILLUSTRATIONS (Cont'd)

6	Variance Ratio $\text{tr}\{\text{var}(\hat{a})\}/\text{tr}[K_a]$ vs Angle Cell Width $\Delta\theta$ and Received Signal-to-Noise Ratio $\rho_N\sigma_a^2$ (Two Angle Cells).....	40
7	Variance Ratio $\text{tr}\{\text{var}(\hat{a})\}/\text{tr}[K_a]$ vs Angle Cell Width $\Delta\theta$ and Received Signal-to-Noise Ratio $\rho_N\sigma_a^2$ (Eight Angle Cells) .....	40
8	Eigenvalue Spread for Narrowband $\Phi$ .....	41
9	Eigenvalue Spread for Wideband $\Phi$ ( $\Delta\theta = 3.62^\circ = \text{BW}/8$ ) .....	43
10	Eigenvalue Spread for Wideband $\Phi$ ( $\Delta\theta = 7.25^\circ = \text{BW}/4$ ) .....	44
11	Eigenvalue Spread for Wideband $\Phi$ ( $\Delta\theta = 10.9^\circ = 3\text{BW}/8$ ).....	45
12	Example of a Two-dimensional Test Volume Geometry; $K = 36$ Test Cells; $M = 1$ Target Cell Under $H_1$ ; $J = 2, 10, 20$ Target Cells Under $H_2$ .....	50
13	Receiver Operating Characteristic (ROC) Curve; $M\beta = J\Psi = 5$ ; $m = 1$ ; $\gamma = \Psi/4$ ; vary $J = 2, 10, 20$ .....	51
14	Nomenclature for Object Motion .....	52
B-1-a	Example of a Two-dimensional Test Volume for Case 1; $K = 36$ Test Cells; $J = 10$ target cells .....	B-8
B-1-b	Example of a Two-dimensional Test Volume for Case 1; $K = 36$ Test Cells; $M = 1$ Target Cell Under $H_1$ ; $J = 10$ Target Cells Under $H_2$ .....	B-9
B-2	Receiver Operating Characteristic (ROC) Curve for Case 1; $J\beta = 10$ , $J\Psi = 2$ , $\beta/\Psi = 5$ , vary $J = 2, 10, 20$ .....	B-12
B-3	$P_{11}$ as a Function of $J$ for Case 1 ( $P_{12} = 0.001$ ) .....	B-13
B-4	Receiver Operating Characteristic (ROC) Curve for Case 1; $J = 20$ ; $\Psi = 0.1$ ; vary $\beta = 0.5, 1, 2$ .....	B-14
B-5	Receiver Operating Characteristic (ROC) Curve for Case 2; $M = 1$ ; $J = 20$ ; $\beta = 5$ ; $\Psi = 0.25$ ; vary $\gamma = 0, 0.05, 0.1, 0.25$ .....	B-18
B-6	Receiver Operating Characteristic (ROC) Curve for Case 2; $M\beta = J\Psi = 5$ ; $M = 1$ ; $\gamma = 0$ ; vary $J = 2, 10, 20$ .....	B-19
B-7	Receiver Operating Characteristic (ROC) Curve for Case 2; $M\beta = J\Psi = 5$ ; $M = 1$ ; $\gamma = \Psi/4$ ; vary $J = 2, 10, 20$ .....	B-20
B-8	Receiver Operating Characteristic (ROC) Curve-Case 2; $M\beta = J\Psi = 10$ ; $J = 20$ ; $\gamma = \Psi$ ; vary $M = 1, 2, 4$ .....	B-21

# A BAYESIAN APPROACH TO ACOUSTIC IMAGING AND OBJECT CLASSIFICATION BY HIGH FREQUENCY SONAR

## INTRODUCTION

Simply stated, the objective of this work is to determine an optimum (in some sense) space-time signal processor for object classification by a high frequency sonar. Only active sonar operating at a frequency and range of the order of 10 kHz (or greater) and several kilometers, respectively, is addressed. Applications of such a sonar are to underwater vehicles, principally weapons and Unmanned Undersea Vehicles (UUVs). The approach is to formulate the classification problem as a statistical decision among several alternative hypotheses, each corresponding to a particular class of true or false target. (The null hypothesis corresponds to only noise in the data.) The criterion for optimality of the test is the Bayes risk, which is to be minimized. This criterion leads to a test based on the likelihood ratio, the ratio of the likelihood functions of the measured data. The data consist of a discrete set of time series representing measurements of the received pressure field at the various positions of the transducer elements in the sonar array. In general, the measurements can be represented as a vector-valued stochastic processes  $x(t)$ , the components corresponding to the measurements at the element positions within the receive array. The process  $x(t)$  is composed of signal and noise, also vector-valued processes. The signal's statistical properties vary among the alternative hypotheses that correspond to the various classes of real and false targets. As shown subsequently, the signal itself is modeled as a randomly weighted sum of time-delayed replicas of the waveform of the transmitted signal. The received signal is then statistically characterized by a multivariate probability density function (pdf) of the random weights. The pdf (its family or, more simply, its moments) differs for each of the various signal (i.e., target) classes. The principal results in this report are derived assuming a Gaussian multivariate pdf; however, the non-Gaussian case as well as more general signal models are discussed.

In summary, the active sonar classification is approached as a likelihood ratio test of multiple, alternative hypotheses versus a noise-only null hypothesis. The data are, in general, vector-valued stochastic processes representing measurements from individual elements within a sonar array. An explicit form is assumed for the received signal model, which is statistically characterized for each

alternative hypothesis (target class). Explicit results are derived for the likelihood ratio, and various performance characteristics are shown. Moreover, the optimal processor is examined from the perspective of acoustic image processing. Generalizations of the results are indicated and in some cases addressed in detail (e.g., the case of moving targets).

## CLASSIFICATION PROBLEM

The sonar *detection* problem is usually formulated as a statistical hypothesis test (references 1 and 2), and the test is binary. Under the null hypothesis  $\mathcal{H}_0$ , the observations or data (i.e., the sensors' outputs) consist of noise alone; under the alternative hypothesis  $\mathcal{H}_1$ , the observations consist of noise and a signal that is the sensors' response to the presence of a target or other interesting object. The sonar *classification* problem can be similarly formulated except that the test is not binary. What distinguishes the classification problem is that there are at least *two* alternative hypotheses because, in addition to the signal produced by the real target, another signal can appear in the observations. In the case of active sonar, this other signal results from the sensors' responses to the pressure field scattered from an object, distinct from the real target; but it resembles in some significant way the signal produced by the real target.

In general, there may be several of these false-target signals each corresponding to a distinct class of scattering objects (either natural or man-made) and each, along with additive noise, representing additional alternative hypotheses regarding the constitution of the data. It follows that the classification problem can be formulated as a multiple hypothesis test with a null hypothesis  $\mathcal{H}_0$  and alternative (exhaustive and mutually exclusive) hypotheses  $\mathcal{H}_i$ , where  $i = 1, 2, \dots, I$ , and  $I > 1$ :

$$\mathcal{H}_0: \mathbf{x}(t) = \mathbf{n}(t), \quad (1a)$$

$$\mathcal{H}_i: \mathbf{x}(t) = \mathbf{s}_i(t) + \mathbf{n}(t), \quad i = 1, 2, \dots, I, \quad (1b)$$

where

$t$  denotes time, and  $T_1 \leq t \leq T_2$ ,  $T_1$  and  $T_2$  being the endpoints of the observation interval,

$\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_N(t)]^T$ , an  $N \times 1$  complex vector,

$x_n(t)$  denotes the output from  $n^{\text{th}}$  sensor element (e.g., the  $n^{\text{th}}$  transducer element in an acoustic array),

$\mathbf{s}_i(t) = [s_{i1}(t), s_{i2}(t), \dots, s_{iN}(t)]^T$ , an  $N \times 1$  complex vector,

$s_{in}(t)$  denotes the signal component in the  $n^{\text{th}}$  sensor output,

$\mathbf{n}(t) = [n_1(t), n_2(t), \dots, n_N(t)]^T$ , an  $N \times 1$  complex vector,

$n_n(t)$  denotes the noise component in the  $n^{\text{th}}$  sensor output,

$n = 1, 2, \dots, N$ , and  $\mathbf{v}^T$  denotes the transpose of any vector  $\mathbf{v}$ .

The vector  $\mathbf{x}(t)$  is a stochastic process that is the output (voltage) of an array of acoustic transducers and comprises the data. Let  $s_1(t)$  be the signal component of  $\mathbf{x}(t)$  corresponding to the



real target;  $s_i(t)$  is the false signal component in  $x(t)$  corresponding to each of the false target hypotheses  $\mathcal{H}_i$ ,  $i = 2, 3, \dots, I$ . The statistical hypothesis test decides among  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_I$  based on measurement over the interval  $[T_1, T_2]$  of  $x(t)$  or, more precisely, realizations of the vector-valued process  $x(t)$ . Note that, in general,  $s_i(t)$  is a stochastic process,  $i = 1, 2, \dots, I$ .

These processes are assumed to be zero-mean, and the signals and noise are uncorrelated; that is,

$$E[n(t)] = 0, \quad (2)$$

$$E[s_i(t)] = 0, \quad i = 1, 2, \dots, I, \quad (3)$$

and thus

$$E[x(t)|\mathcal{H}_i] = 0, \quad i = 0, 1, \dots, I, \quad (4)$$

and

$$E[n(t_1) s_i^\dagger(t_2)] = 0, \quad (5)$$

for all  $t, t_1, t_2$  in  $[T_1, T_2]$  ( $v^\dagger$  denotes the complex conjugate-transpose of any vector  $v$ ). The  $N \times N$  covariance matrices  $K_n(t_1, t_2)$  and  $K_{s_i}(t_1, t_2)$  are defined by

$$K_n(t_1, t_2) = E[n(t_1)n^\dagger(t_2)], \quad (6)$$

$$K_{s_i}(t_1, t_2) = E[s_i(t_1)s_i^\dagger(t_2)], \quad i = 1, 2, \dots, I, \quad (7)$$

and

$$K_{x_i}(t_1, t_2) = E[x(t_1)x^\dagger(t_2)|\mathcal{H}_i], \quad i = 0, 1, \dots, I. \quad (8)$$

for all  $t_1, t_2$  in  $[T_1, T_2]$ . Then, from equations (1) and (5),

$$K_{x_i}(t_1, t_2) = K_{s_i}(t_1, t_2) + K_n(t_1, t_2), \quad i = 1, 2, \dots, I. \quad (9)$$

and

$$\mathbf{K}_{x_0}(t_1, t_2) = \mathbf{K}_n(t_1, t_2),$$

for all  $t_1, t_2$  in  $[T_1, T_2]$ . Note that, from equation (8),

$$\mathbf{K}_{x_i}^\dagger(t_1, t_2) = \mathbf{K}_{x_i}(t_2, t_1), \quad (10)$$

for all  $t_1, t_2$  in  $[T_1, T_2]$  and  $i = 0, 1, 2, \dots, I$ .

For the binary case ( $I = 1$ ), the optimal (Bayesian) test for vector-valued processes ( $N > 1$ ) is derived in references 3 and 4 based on the vector-valued Karhunen-Loève expansion (reference 5). The optimal (minimum probability of error) test in the general case of multiple alternative hypotheses ( $I > 2$ ) is given in reference 6, but is restricted to the scalar case ( $N = 1$ ) and is not directly applicable to multiple channels (sensors). Multiple alternatives and vector-valued data are treated in reference 7, but only when the signal vector is assumed known *a priori* for all alternatives. In a treatment of diversity reception (reference 8), a similar case is addressed wherein a multiplicative, random disturbance is present along with additive noise; however, the multiplicative noise is identical under all alternative hypotheses. Hence, such a model does not apply to the classification problem as formulated here. When the signal and noise processes are *stationary*, then, of course, the covariance matrix, of the data  $\mathbf{x}(t)$  can be written as  $\mathbf{K}_{x_i}(t_1, t_2) = \mathbf{K}_{x_i}(t_1 - t_2)$ , for all  $t_1, t_2$  in  $[T_1, T_2]$  and  $i = 0, 1, 2, \dots, I$ . In this case the analysis proceeds using the spectral density matrix which exists as the Fourier transform of  $\mathbf{K}_{x_i}(t_1 - t_2)$ . This is well-covered in the literature; references 9 - 13 are representative. The signals and noise received by passive sonar can usually be assumed to be stationary. However, when active sonar is used, the signals and noise (reverberation) are inherently nonstationary; they are received as the pressure field backscattered from complex objects (i.e., objects with nonuniform scattering strength). Therefore, it is the inherent nonstationarity of the signals that is the basis for the discrimination. To assume stationarity would be to overlook a characteristic feature of the signal that reveals the spatial character of a scattering object.

In addition to reference 6 cited previously, there are theoretical treatments of the general classification problem, references 14 and 15, for example. While these address the case of  $I \geq 2$ , they typically consider only the case of a scalar process ( $N = 1$ ). Moreover, they do not introduce models of the signal processes (other than the Gaussian assumption). Signals must be

modeled to infuse into the classifier their distinguishing *physical* characteristics. In the next section, a simple but general model of the signal processes is introduced; it applies to active sonar.

In summary, the unique contribution presented herein is the systematic integration of all of the essential features of the active sonar classification process: (1) multiple alternative hypotheses ( $I \geq 2$ ); (2) multiple channels (sensors) represented by vector-valued processes ( $N \geq 2$ ); (3) a stochastic, but not necessarily stationary, model of the alternative signal processes that applies to the active sonar problem; (4) an optimal (Bayesian) decision rule where the costs of errors (incorrect decisions) are not necessarily identical. Moreover, a concise expression for the optimal processor is derived.

## SIGNAL MODEL

The  $i^{\text{th}}$  signal vector is, as in equation (1), given by

$$\mathbf{s}_i(t) = [s_{i1}(t), s_{i2}(t), \dots, s_{iN}(t)]^T, \quad i = 1, 2, \dots, I.$$

Let the  $n^{\text{th}}$  component of  $\mathbf{s}_i(t)$  be represented as a sum of time-delayed and amplitude-weighted replicas of the transmitted waveform  $f(t)$ ; i.e., let, for  $n = 1, 2, \dots, N$ ,

$$s_{in}(t) = \sum_{k=1}^{K_i} a_{ik} f(t - \tau_{ink}), \quad T_1 \leq t \leq T_2, \quad (11)$$

where, for each  $i = 1, 2, \dots, I$ ,  $\{a_{ik}\}_1^{K_i}$  is a sequence of complex random variables with the following properties:

$$E(a_{ik}) = 0, \quad (12)$$

and

$$E(a_{ij} a_{kl}^*) = [\mathbf{K}_{a_i}]_{jl} \delta_{ik}, \quad (13)$$

where  $j, k, l = 1, 2, \dots, K_i$ ;  $\mathbf{K}_{a_i}$  is the  $(K_i \times K_i)$  covariance matrix of the random variables  $\{a_{ik}\}_1^{K_i}$ ,  $i = 1, 2, \dots, I$ ; and  $\delta_{ik}$  is the Kronecker delta function.

If it is assumed that a given object — the  $i^{\text{th}}$  one — is a set of  $K_i$  distinct stationary scatterers, then the signal model, equation (11), is interpreted as follows. The index  $k$  is over the set of scatterers: each value of  $k$  corresponds to one and only one scatterer,  $a_{ik}$  is its scattering coefficient;  $\tau_{ink}$  is the delay from the instant ( $t = 0$ ) of transmission (from a reference transducer element in the array) to reception by the  $n^{\text{th}}$  element of the array of the waveform scattered by the  $k^{\text{th}}$  scatterer. Under this interpretation, analogous to the so-called Lagrangian (or material) description in fluid mechanics,  $K_i$  and the delays  $\{\tau_{ink}\}_1^{K_i}$  are not fixed at known values within a reference coordinate system. In general, they would be treated as random variables (at least for naturally occurring false targets); thus, specification of their joint distribution function is required.

Moreover,  $K_i$  must, in general, be assumed to be a random variable having a discrete distribution such as the Poisson, for example. This random model obtains from the empirical observation that natural objects that appear as false targets are, at sufficiently high frequencies, ensembles of reflecting facets whose spatial distributions are random. Since determination of realistic probability distribution functions of the appropriate random variables is difficult, the Lagrangian description is problematic and will not be further considered.

A more suitable alternative is the Eulerian (or spatial) description. Following this description,  $a_{ik}$  is interpreted as the random scattering coefficient corresponding to a specific volume element fixed within a given (three-dimensional) coordinate system. The origin of the coordinate system is conveniently fixed at some reference point, a transducer element in the receiving array, for example, as in figure 1. The figure illustrates an example with spherical coordinates and a uniform planar array in a nonrefractive medium. The set of volume elements defines a region in the medium, a so-called test region, for which one and only one of  $\{\mathcal{H}_i\}_0^I$  is true. More precisely, the test region is defined as a set of disjoint volume elements called cells that, it is assumed, contain at most one scatterer. Otherwise, the cell size remains, for now, unspecified. When the test region is occupied by the  $i^{\text{th}}$  object, the scattering from the region is characterized by the set of coefficients  $\{a_{ik}\}_1^{K_i}$ , for  $i = 1, 2, \dots, I$ . The delay  $\tau_{ink}$  is now the delay from  $t = 0$  to reception at the  $n^{\text{th}}$  transducer element of the waveform scattered from the  $k^{\text{th}}$  cell (or, more precisely, that part of the  $i^{\text{th}}$  object that occupies the  $k^{\text{th}}$  cell). It can be assumed that, for each  $\mathcal{H}_i$ ;  $i = 1, 2, \dots, I$ , the test region and its decomposition are specified *a priori*. Thus, the number of cells  $K_i$  and the position of each cell (and, hence, the values of the delays  $\{\tau_{ink}\}_1^{K_i}$ ) are specified constants for all  $i = 1, 2, \dots, I$ . Of course, as search for the actual target progresses, the test region will shift in some prescribed way through the insonified volume during each transmit cycle. The values of  $\{\tau_{ink}\}_1^{K_i}$  and  $K_i$  will change accordingly.

By adopting the Eulerian description, the randomness of the signal is characterized in the model solely by the statistical properties of the coefficients  $\{a_{ik}\}_1^{K_i}$ , which, to second order, are specified by the covariance matrix  $\mathbf{K}_{a_i}$ ,  $i = 1, 2, \dots, I$ . In what follows,  $\mathbf{K}_{a_i}$  is not required to be of full rank; therefore, the model allows for coherence among the waveforms scattered from each cell. In practice, the dimensions and decomposition of test regions and estimates of  $\mathbf{K}_{a_i}$  will be determined directly from both experimental data and calculations from scattering models of the hypothesized targets. Of course, the dimensions of the test regions would match, roughly at least, the characteristic dimensions of the targets.

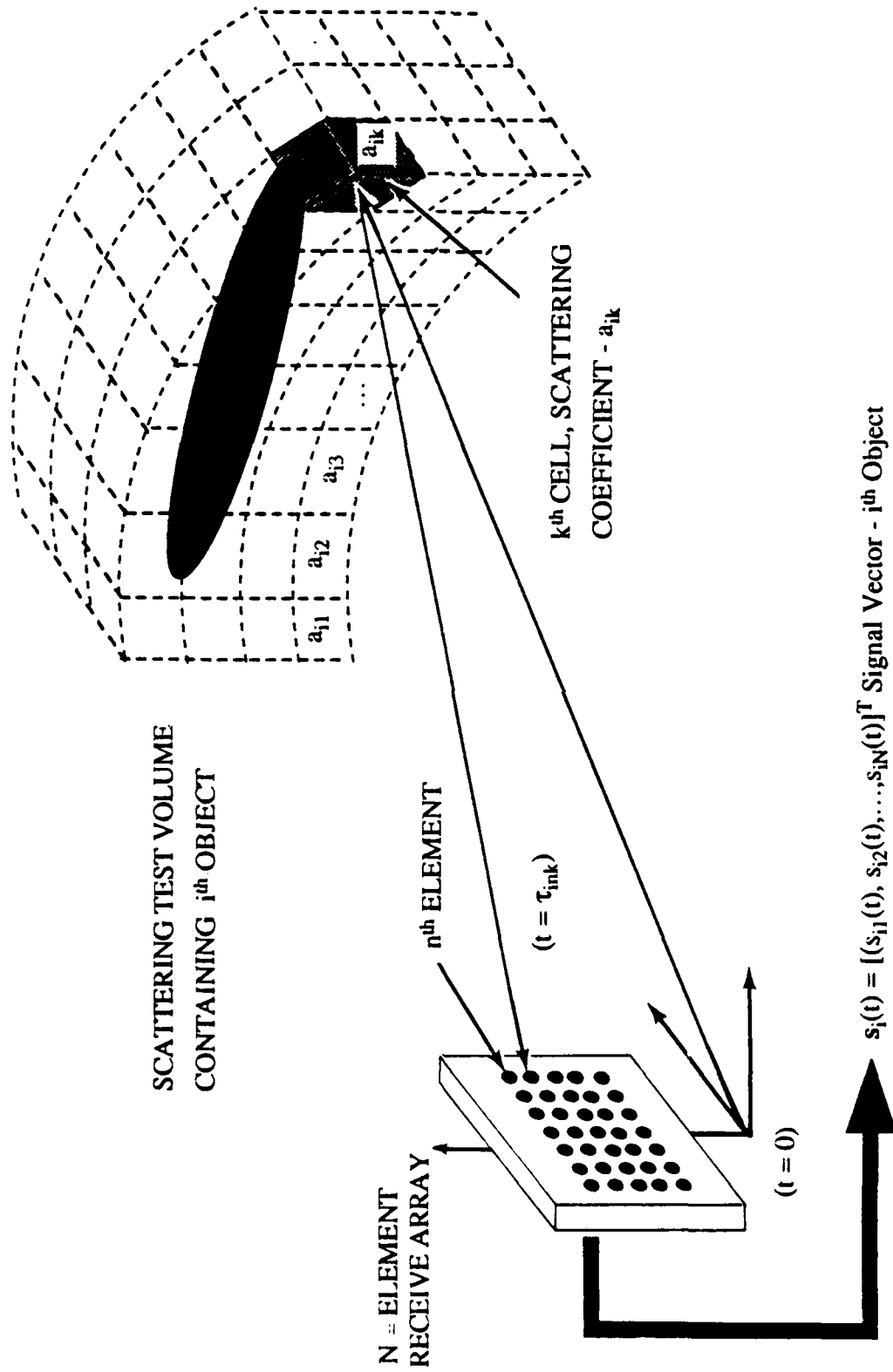


Figure 1. Signal Model

To summarize, equation (11) is rewritten in vector form:

$$s_i(t) = F_i(t)a_i, \quad T_1 \leq t \leq T_2, \quad i = 1, 2, \dots, I, \quad (14)$$

where  $s_i(t)$  is as in equation (1);  $a_i = [a_{i1}, a_{i2}, \dots, a_{iK_i}]^T$ , a  $K_i \times 1$  complex vector;  $F_i(t)$  is an  $N \times K_i$  complex matrix with elements given by  $[F_i(t)]_{nk} = f(t - \tau_{ink})$ ,  $n = 1, 2, \dots, N$  and  $k = 1, 2, \dots, K_i$ . The following assumptions pertain to equation (14):

1. Implicit in equation (14) is that scattering from a given cell is frequency independent at least over the band of  $f(t)$ ; otherwise, the summand on the right side of equation (14) would be a temporal convolution.
2. The test region comprises  $K_i$  disjoint volume elements (cells) that contain, at most, one scattering facet.
3.  $f(t)$ , the transmit waveform, is known and has finite energy  $\int_{-\infty}^{\infty} |f(t)|^2 dt = E_f$ ; otherwise,  $f(t)$  is arbitrary, as consistent with assumption 1.
4.  $E(a_i) = 0$ ,  $i = 1, 2, \dots, I$ , as previously indicated.
5. Neither the scatterers nor the sonar platform is in motion (this assumption will be relaxed later).

It follows from equation (14) that the covariance matrix of  $s_i(t)$  is given by

$$K_{s_i}(t_1, t_2) = F_i(t_1)K_{a_i}F_i^\dagger(t_2), \quad (15)$$

and

$$E[s_i(t)] = 0, \quad i = 1, 2, \dots, I, \quad (16)$$

reiterating equation (3). Furthermore, from equation (15), the trace of  $K_{s_i}(t, t)$  is given by, for  $i = 1, 2, \dots, I$ ,

$$\text{tr}[K_{s_i}(t, t)] = \text{tr}[F_i^\dagger(t)F_i(t)K_{a_i}], \quad (17)$$

and, therefore, the total energy in the received signal is given by

$$\int_{T_2}^{T_1} \text{tr}[\mathbf{K}_{s_i}(t,t)]dt = NE_f \text{tr}[\Phi_i \mathbf{K}_{a_i}], \quad (18)$$

where  $\Phi_i$  is the positive definite, Hermitian ( $K_i \times K_i$ ) matrix,

$$\Phi_i = \frac{1}{NE_f} \int_{T_1}^{T_2} \mathbf{F}_i^\dagger(t) \mathbf{F}_i(t) dt, \quad (19)$$

and is the normalized waveform correlation matrix. If, for example, the scattering is incoherent from cell to cell, then the matrix  $\mathbf{K}_{a_i}$  is diagonal and equation (18) becomes

$$\int_{T_2}^{T_1} \text{tr}[\mathbf{K}_{s_i}(t,t)]dt = NE_f \text{tr}[\mathbf{K}_{a_i}], \quad (20)$$

as expected.

Note that this kind of signal model is more flexible than it might first appear. The matrix  $\mathbf{K}_{a_i}$  can account for coherence among the signals scattering from the cells in the test region. As mentioned above, incoherent scattering is characterized by letting  $\mathbf{K}_{a_i}$  be diagonal (but not necessarily of full rank, since some diagonal terms may be identically zero when the corresponding cell is empty). When coherent scattering occurs,  $\mathbf{K}_{a_i}$  will have off-diagonal terms accounting for correlations between pairs from  $\{a_{ik}\}_1^{K_i}$  corresponding usually to neighboring cells.

## NOISE MODEL

The zero-mean noise process  $\mathbf{n}(t)$  in equation (1) has covariance matrix  $\mathbf{K}_n(t_1, t_2)$  such that

$$\int_{T_1}^{T_2} \text{tr}[\mathbf{K}_n(t,t)]dt < \infty. \quad (21)$$



It is assumed that  $\mathbf{n}(t)$  always contains a white noise component; the colored noise component has a covariance matrix with elements that are uniformly bounded and continuous for all  $t_1, t_2$ , in  $[T_1, T_2]$ .

Furthermore, it is assumed that for the noise process  $\mathbf{n}(t)$  there exists a *reversible* whitening filter with an impulse response ( $N \times N$ ) matrix  $\mathbf{W}(t_1, t_2)$ ; so that if  $\mathbf{w}(t)$  is a vector-valued, white noise process, then

$$\mathbf{w}(t) = \int_{T_1}^{T_2} \mathbf{W}(t, \tau) \mathbf{n}(\tau) d\tau, \quad T_1 \leq t \leq T_2. \quad (22)$$

which leads to

$$\frac{N_0}{2} \delta(t_1 - t_2) \mathbf{I} = \int_{T_1}^{T_2} \int_{T_1}^{T_2} \mathbf{W}(t_1, \tau) \mathbf{K}_n(\tau, \tau') \mathbf{W}^+(\tau', t_2) d\tau d\tau' \quad (23)$$

for  $t_1, t_2$  in  $[T_1, T_2]$ ;  $N_0/2$  is the spectral density level of the white noise;  $\mathbf{I}$  is the  $N \times N$  identity matrix. Define  $\mathbf{y}(t)$  by

$$\mathbf{y}(t) = \int_{T_1}^{T_2} \mathbf{W}(t, \tau) \mathbf{x}(\tau) d\tau, \quad T_1 \leq t \leq T_2; \quad (24)$$

then the model given by equations (1) and (14) is transformed to

$$\mathcal{H}_0: \mathbf{y}(t) = \mathbf{w}(t), \quad (25a)$$

$$\mathcal{H}_i: \mathbf{y}(t) = \mathbf{G}_i(t) \mathbf{a}_i + \mathbf{w}(t), \quad (25b)$$

where  $\mathbf{G}_i(t)$  is the  $N \times K_i$  matrix given by

$$\mathbf{G}_i(t) = \int_{T_1}^{T_2} \mathbf{W}(t, \tau) \mathbf{F}_i(\tau) d\tau, \quad (26)$$

for  $i = 1, 2, \dots, I$  and  $T_1 \leq t \leq T_2$ . Equations (25) are of the same form as equations (1) and (14):  $G_i(t)$  replaces  $F_i(t)$ , both being the known, matrix-valued functions that pre-multiply the random vector  $a_i$ . Because  $W(t_1, t_2)$  is reversible, it can be applied to  $x(t)$ , as in equation (24), to facilitate the solution to the optimal decision problem, i.e., the optimal hypotheses test given the data,  $x(t)$ . The reversibility of  $W(t_1, t_2)$  is defined as the existence of an inverse  $W^{-1}(t_1, t_2)$  such that

$$\int_{T_1}^{T_2} W^{-1}(t_1, \tau) W(\tau, t_2) d\tau = \delta(t_1 - t_2) I, \quad (27)$$

for  $t_1, t_2$  both in  $[T_1, T_2]$ . The reversibility of  $W(t_1, t_2)$  is a sufficient condition for the equivalence of equations (25) and equations (1) and (14) for both the decision problem *and* the estimation problem (reference 1). The estimation problem will be discussed briefly in the next section. In what follows, it is assumed that the noise process is white and that the form given by equations (1) and (14) holds, understanding that an identical form is obtained if the noise is colored and a (reversible) prewhitening filter is applied as explained.

## ESTIMATION OF $a_i$

Before developing the optimal hypothesis test (classifier), it is instructive to first determine the optimal estimator of the random vector  $a_i$  from the data  $x(t)$ , which under  $\mathcal{H}_i$  ( $i = 1, 2, \dots, I$ ) are given by the linear model (combining equations (1b) and (14))

$$x(t) = F_i(t) a_i + n(t), \quad T_1 \leq t \leq T_2. \quad (28)$$

An estimator  $\hat{a}_i$  is sought that, under  $\mathcal{H}_i$ , is a minimum variance estimator of  $a_i$  and that is unbiased and linear as follows:

$$E(\hat{a}_i | \mathcal{H}_i) = E(a_i), \quad (29)$$

and

$$\hat{a}_i = \int_{T_1}^{T_2} H_i(t) x(t) dt, \quad i = 1, 2, \dots, I, \quad (30)$$

where  $H_i(t)$  is a  $K_i \times N$  matrix function to be determined. Note that  $E(a_i) = 0$  and  $E[n(t)] = 0$ ; therefore,  $E[x(t)|\mathcal{H}_i] = 0$ ; and, thus,  $E(\hat{a}_i) = 0$ , for  $i = 1, 2, \dots, I$ . Let  $\epsilon_i$  represent the estimation error,  $\epsilon_i = \tilde{a}_i - a_i$ , and consider the quadratic form  $q(\tilde{a})$  given by

$$q(\tilde{a}_i) = E(\epsilon_i^\dagger V \epsilon_i | \mathcal{H}_i),$$

where  $V$  is any  $K_i \times K_i$  positive definite matrix and  $\tilde{a}_i$  is any estimator that satisfies equations (29) and (30). Then  $q(\tilde{a}_i)$  is minimized when  $\tilde{a}_i = \hat{a}_i$  and  $H_i(t)$  satisfies the special form of the Wiener-Hopf equation (references 18 and 19) given by

$$E[a_i x^\dagger(t) | \mathcal{H}_i] = \int_{T_2}^{T_1} H_i(\tau) E[x(\tau) x^\dagger(t) | \mathcal{H}_i] d\tau, \quad (31)$$

for  $i = 1, 2, \dots, I$  and  $T_1 \leq t \leq T_2$ . From equation (28),

$$E[a_i x^\dagger(t) | \mathcal{H}_i] = K_{a_i} F_i^\dagger(t), \quad i = 1, 2, \dots, I, \quad (32)$$

since  $E[a_i n^\dagger(t)] = 0$  for all  $t$  in  $[T_1, T_2]$ . From equations (9) and (15),

$$E[x(t_1) x^\dagger(t_2) | \mathcal{H}_i] = F_i(t_1) K_{a_i} F_i^\dagger(t_2) + \frac{N_0}{2} \delta(t_1 - t_2) I, \quad i = 1, 2, \dots, I, \quad (33)$$

where, as explained before, it has been assumed that

$$K_n(t_1, t_2) = \frac{N_0}{2} \delta(t_1 - t_2) I, \quad T_1 \leq t_1, t_2 \leq T_2.$$

Equation (31) becomes

$$K_{a_i} F_i^\dagger(t) = \int_{T_2}^{T_1} H_i(\tau) F_i(\tau) d\tau K_{a_i} F_i^\dagger(t) + \frac{N_0}{2} H_i(t), \quad (34)$$

for  $i = 1, 2, \dots, I$  and  $T_1 \leq t \leq T_2$ . Now, let  $H_i(t) = \frac{2}{N_0} H_i F_i^\dagger(t)$ , where  $H_i$  is a  $K_i \times K_i$  matrix to be determined, and substitute into equation (34). This yields, for  $i = 1, 2, \dots, I$ , the solution for  $H_i$  given by

$$H_i = K_{a_i} \left[ I + \frac{2NE_f}{N_0} \Phi_i K_{a_i} \right]^{-1}, \quad (35)$$

where as before  $\Phi_i$  is the normalized waveform correlation ( $K_i \times K_i$ ) matrix given by

$$\Phi_i = \frac{1}{NE_f} \int_{T_1}^{T_2} F_i^\dagger(t) F_i(t) dt. \quad (36)$$

Note that equation (34) is a matrix form of the Fredholm integral equation of the second kind. A unique solution exists since, in this case, the components of the matrix  $K_{s_i}(t_1, t_2) = F_i(t_1) K_{a_i} F_i^\dagger(t_2)$  are continuous in  $t_1, t_2$  over  $[T_1, T_2]$  and are uniformly bounded. The optimal (under  $\mathcal{H}_i$ ) estimator  $\hat{a}_i$  is then given by (from equations (30) and (35)):

$$\hat{a}_i = \frac{2}{N_0} K_{a_i} \left[ I + \rho_N \Phi_i K_{a_i} \right]^{-1} \int_{T_1}^{T_2} F_i^\dagger(t) x(t) dt, \quad i = 1, 2, \dots, I, \quad (37)$$

where  $\rho_N = \frac{2NE_f}{N_0}$ .

It can be shown, using equation (37), that the estimation error covariance matrix is given by

$$E[(\hat{a}_i - a_i)(\hat{a}_i - a_i)^\dagger | \mathcal{H}_i] = K_{a_i} \left[ I + \rho_N \Phi_i K_{a_i} \right]^{-1}, \quad i = 1, 2, \dots, I. \quad (38)$$

Note that this estimation error covariance is conditional on  $\mathcal{H}_i$  being true. Of course, the estimator  $\hat{a}_i$  given by equation (37) applies only when  $\mathcal{H}_i$  is true; its applicability will become apparent when the optimal classifier is developed in subsequent sections.

Equation (38) shows explicitly the dependence of the estimation error covariance on *a priori* information regarding the  $i^{\text{th}}$  scattering object's spatial properties. Moreover, the dependence on

system characteristics — the array geometry and the transmitted waveform — is explicitly represented by the matrix  $\Phi_i$ . Thus, the effect of alternative array and waveform designs can be evaluated directly through equation (38). For example, the performance of sparse (two-dimensional) arrays of receive transducers can be assessed for acoustic imaging applications.

## OPTIMAL BAYES CLASSIFIER

The optimal classifier considered here is a test of the multiple hypotheses  $\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_I$  that is optimal according to the Bayes criterion. An optimal Bayes test is one that minimizes the risk (i.e., the expected loss) in applying a given decision rule. Theoretical details and proof can be found in a variety of texts (references 1, 20, and 21). The Bayesian risk  $\mathcal{R}$  is defined as (reference 20)

$$\mathcal{R} = \sum_{i=0}^I \sum_{j=0}^I c_{ij} p_j \Pr[\mathcal{H}_i | \mathcal{H}_j], \quad (39)$$

where  $c_{ij}$  is the cost of deciding hypothesis  $\mathcal{H}_i$  when  $\mathcal{H}_j$  is true;  $p_j$  is  $\Pr[\mathcal{H}_j]$ , the *a priori* probability of  $\mathcal{H}_j$ ;  $\Pr[\mathcal{H}_i | \mathcal{H}_j]$  is the probability that a given decision rule chooses  $\mathcal{H}_i$  when  $\mathcal{H}_j$  is true;  $i, j = 0, 1, 2, \dots, I$ . Note that

$$\sum_{j=0}^I p_j = 1,$$

and

$$\Pr[\mathcal{H}_i | \mathcal{H}_j] = \int_{Z_i} p(\mathbf{x} | \mathcal{H}_j) d\mathbf{x}, \quad (40)$$

where  $p(\mathbf{x} | \mathcal{H}_j)$  is the conditional probability density function of a measurement vector  $\mathbf{x}$  given  $\mathcal{H}_j$ ;  $Z_i$  is a region in the sample space of  $\mathbf{x}$  such that  $\mathcal{H}_i$  is chosen if a realization of  $\mathbf{x}$  is contained in  $Z_i$ ; the regions  $\{Z_i\}_0^I$  are disjoint and cover the entire sample space. Here  $\mathbf{x}$  instead of  $\mathbf{x}(t)$  is used for reasons given in the next section. It can be shown (reference 20) that the risk  $\mathcal{R}$  is minimized by the following decision rule:

$$\text{choose } \mathcal{H}_k \text{ if } D_k(\mathbf{x}) = \max\{D_0(\mathbf{x}), D_1(\mathbf{x}), \dots, D_I(\mathbf{x})\}, \quad (41)$$

where  $D_i$  is the so-called *discriminant score* defined as

$$D_i(\mathbf{x}) = -\sum_{j=0}^I c_{ij} p_j p(\mathbf{x}|\mathcal{H}_j), \quad i = 0, 1, 2, \dots, I. \quad (42)$$

Note that for the special case of  $c_{ij} = 1 - \delta_{ij}$ ,  $\mathcal{R}$  is the total probability of error and  $D_i$  reduces to  $D_i(\mathbf{x}) = p_i p(\mathbf{x}|\mathcal{H}_i) + c$ ,  $i = 0, 1, 2, \dots, I$ , where  $c$  is a constant independent of  $i$ . In this special case,  $D_i(\mathbf{x})$  is directly proportional to  $\Pr[\mathcal{H}_i|\mathbf{x}]$ , the *a posteriori* probability of the hypothesis  $\mathcal{H}_i$ , since

$$\Pr[\mathcal{H}_i|\mathbf{x}] = \frac{p(\mathbf{x}|\mathcal{H}_i) p_i}{\sum_{i=0}^I p(\mathbf{x}|\mathcal{H}_i) p_i}. \quad (43)$$

The decision rule defined by equations (41) and (43) is the so-called maximum *a posteriori* decision rule. In this report, the decision rule is not restricted to this special case, but preserved in the more general form of the decision rule of equation (42).

Now, observe that equation (42) can be written as follows:

$$D_i(\mathbf{x}) = p(\mathbf{x}|\mathcal{H}_0) \sum_{j=1}^I \lambda_{ij} \Lambda_j(\mathbf{x}), \quad (44)$$

where

$$\Lambda_j(\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{H}_j)}{p(\mathbf{x}|\mathcal{H}_0)},$$

which is the likelihood ratio of  $\mathcal{H}_j$  with respect to  $\mathcal{H}_0$  for  $j = 1, 2, \dots, I$ . Note that  $\Lambda_0 = 1$  and  $\lambda_{ij}$  is defined by

$$\lambda_{ij} = -c_{ij} p_j, \quad i, j = 0, 1, 2, \dots, I.$$

Since the factor  $p(\mathbf{x}|\mathcal{H}_0)$  is common to all  $D_i(\mathbf{x})$ , an equivalent decision rule is given by

$$\text{choose } \mathcal{H}_k \text{ if } E_k(\mathbf{x}) = \max\{E_0(\mathbf{x}), E_1(\mathbf{x}), \dots, E_I(\mathbf{x})\}, \quad (45)$$

where  $E_i(\mathbf{x})$  is defined by

$$E_i(\mathbf{x}) = \sum_{j=0}^I \lambda_{ij} \Lambda_j(\mathbf{x}), \quad i = 0, 1, \dots, I. \quad (46)$$

Note that while the decision statistic  $E_i(\mathbf{x})$  is a linear function of the likelihood ratios  $\{\Lambda_i(\mathbf{x})\}_0^I$  it is not in general a linear function of the data  $\mathbf{x}$ . The functional dependency on  $\mathbf{x}$  obtains, of course, from the specific distributional form assumed for the likelihood functions  $p(\mathbf{x}|\mathcal{H}_i)$ ,  $i = 0, 1, 2, \dots, I$ . The linear dependency of  $E_i(\mathbf{x})$  on the set of constants  $\{\lambda_{ij}\}_0^I$ , which in this report are called classifier weights, is convenient. Of course, strict adherence to the Bayesian formalism requires values for each cost  $c_{ij}$  and each *a priori* probability  $p_j$  to compute the corresponding weight  $\lambda_{ij}$ . Specification of these values is usually impractical; although, assuming equal *a priori* probabilities, i.e., assuming  $p_j = 1/(I+1)$ ,  $j = 0, 1, 2, \dots, I$ , may be reasonable. Moreover, the costs (however "cost" may be defined in this context) of correct decisions  $\{c_{ij}\}_0^I$  may be reasonably set to zero. Given these assumptions,

$$\lambda_{ij} = \begin{cases} 0 & i = j \\ \frac{c_{ij}}{I+1} & i \neq j \end{cases} \quad i, j = 0, 1, \dots, I. \quad (47)$$

Further specification of the costs is difficult given that a meaningful cost measure (such as time spent searching by the sonar platform) can even be defined in the first place. It would not be reasonable to assume, as in the case that leads to the maximum *a posteriori* decision rule, that all errors have identical costs. For example, consider the case  $I = 2$ . Let  $\mathcal{H}_0$  denote noise only,  $\mathcal{H}_1$  denote *real* target and noise, and  $\mathcal{H}_2$  denote *false* target and noise; then perhaps  $c_{02} = c_{20}$ , but certainly  $c_{02} < c_{12}$  or  $c_{21}$ , for example.

These difficulties become compounded as  $I$  increases, since the number of error types is  $I(I+1)$ . Nonetheless, the Bayesian decision rule given by equations (44) and (45) does give structure to classifier design and shows how the likelihood ratios  $\{\Lambda_i(\mathbf{x})\}_0^I$  figure in the design. This structure, when applied to experimental data, offers a method of determining the coefficients  $\{\lambda_{ij}\}$  empirically (given auxiliary constraints like equation (47), for example). One possible approach is to: (1) specify values of the probabilities of each type of error (except one); (2)

determine from experimental data empirical operating characteristics curves by iterative variation of the coefficients  $\{\lambda_{ij}\}$ , selecting values corresponding to the assigned error probabilities.

A theoretical approach to determining the coefficients  $\{\lambda_{ij}\}$  is to apply a generalization of the Neyman-Pearson Lemma (references 7,20, and 21) to the multiple hypothesis test directly instead of using the Bayes criterion. However, this approach leads to a decision rule equivalent to that specified by equations (45) and (46) except that the coefficients  $\{\lambda_{ij}\}$  could, in principle, be expressed directly as functions of the specified error probabilities.

## LIKELIHOOD RATIO

### REPRESENTATION OF $x(t)$

Before proceeding further, the technical problems of a discrete representation of the vector-valued process  $x(t)$  must be addressed. This will facilitate development of the optimal Bayes test as will be shown in the next section. An expansion is sought of the form

$$x(t) = \text{l.i.m.}_{M \rightarrow \infty} \sum_{m=1}^M x_m \Psi_m(t), \quad (48)$$

where  $\{x_m\}$  are (scalar) random variables,  $\{\Psi_m(t)\}$  is a set of deterministic, orthonormal  $N \times 1$  vector-valued functions (complete over  $C^N x[T_1, T_2]$ ), and "l.i.m." denotes limit in the mean square sense. Equation (48) can be restated as

$$x(t) = \text{l.i.m.}_{M \rightarrow \infty} x_M(t), \quad (49)$$

where  $x_M(t)$  is defined by

$$x_M(t) = \sum_{m=1}^M x_m \Psi_m(t). \quad (50)$$

Since  $\{\Psi_m(t)\}$  is a set of orthonormal functions, then



$$x_m = \int_{T_1}^{T_2} \Psi_m^\dagger(t) x_M(t) dt, \quad m = 1, 2, \dots, M. \quad (51)$$

First consider  $n(t)$ , a white noise process. It can be shown (given methods in reference 5) that its Karhunen-Loève expansion can be made with any complete set of orthonormal functions; i.e.,

$$n(t) = \lim_{M \rightarrow \infty} \sum_{m=1}^M n_m \Psi_m(t), \quad (52)$$

where

$$n_m = \int_{T_1}^{T_2} \Psi_m^\dagger(t) n(t) dt, \quad (53)$$

from which it follows that, for all  $m, n = 1, 2, \dots$ ,

$$E[n_m^* n_n] = \frac{N_o}{2} \delta_{mn}, \quad (54)$$

since  $K_n(t_1, t_2) = (N_o/2) \delta(t_1 - t_2) \mathbf{I}$  for all  $t_1, t_2$  in  $[T_1, T_2]$ .

The next step is to expand  $s_i(t)$  using the set  $\{\Psi_m(t)\}$ ; i.e., expand  $s_i(t)$  as

$$s_i(t) = \lim_{M \rightarrow \infty} \sum_{m=1}^M s_{im} \Psi_m(t), \quad i = 0, 1, 2, \dots, I.$$

An expansion is sought such that only the random sequence  $\{s_{im}\}$  depends on  $i$ , where, as before, the index denotes the hypothesis  $\mathcal{H}_i$ . Note that the set  $\{\Psi_m(t)\}$  is to be independent of  $i$ , and orthogonality leads to

$$s_{im} = \int_{T_1}^{T_2} \Psi_m^\dagger(t) s_i(t) dt, \quad m = 1, 2, \dots \quad (55)$$

Here it is assumed that  $s_i(t)$  is of the form given by equation (14)

$$s_i(t) = F_i(t)a_i,$$

where  $F_i(t)$  is a deterministic  $(N \times K_i)$  matrix, and  $a_i$  is a random  $(K \times 1)$  vector. The matrix  $F_i(t)$  can be expressed as

$$F_i(t) = [f_{i1}(t), f_{i2}(t), \dots, f_{iK_i}(t)], \quad (56)$$

where  $f_{ik}(t)$  is an  $(N \times 1)$  vector that is deterministic and can, therefore, be expanded by the complete set  $\{\Psi_m(t)\}$ :

$$f_{ik}(t) = \lim_{M \rightarrow \infty} \sum_{m=1}^M f_{ikm} \Psi_m(t), \quad k = 1, 2, \dots, K_i, \quad (57)$$

and

$$f_{ikm} = \int_{T_1}^{T_2} \Psi_m^\dagger(t) f_{ik}(t) dt, \quad m = 1, 2, \dots \quad (58)$$

By writing equation (14) as

$$s_i(t) = \sum_{k=1}^{K_i} a_{ik} f_{ik}(t), \quad (59)$$

$s_i(t)$  can then be expressed, using equation (57), as

$$s_i(t) = \lim_{M \rightarrow \infty} \sum_{m=1}^M \left( \sum_{k=1}^{K_i} a_{ik} f_{ikm} \right) \Psi_m(t), \quad (60)$$

which is the desired expansion with

$$s_{im} = \sum_{k=1}^{K_i} a_{ik} f_{ikm}, \quad m = 1, 2, \dots; \quad i = 1, 2, \dots, I. \quad (61)$$

Now,  $x(t)$  can be expanded as in equation (48) where, under  $\mathcal{H}_i$ , the sequence  $\{x_m\}$  is given by

$$x_m = s_{im} + n_m, \quad m = 1, 2, \dots, \quad (62)$$

with the sequences  $\{s_{im}\}$  and  $\{n_m\}$  being defined by equations (61) and (53), respectively. If the vector  $\mathbf{x}$  is defined as  $\mathbf{x} = [x_1, x_2, \dots, x_M]^T$ , then, under  $\mathcal{H}_i$ ,

$$\mathbf{x} = \mathbf{s}_i + \mathbf{n}, \quad (63)$$

where  $\mathbf{s}_i$  and  $\mathbf{n}$  are similarly defined vectors. Then, the following hold:

$$E(\mathbf{x} | \mathcal{H}_i) = \mathbf{0}, \quad (64)$$

since  $s_i(t)$  and  $\mathbf{n}(t)$  are zero-mean processes;

$$E(\mathbf{x} \mathbf{x}^\dagger | \mathcal{H}_i) = E(s_i s_i^\dagger) + (N_0/2) \mathbf{I}, \quad (65)$$

from the independence of  $s_i(t)$  and  $\mathbf{n}(t)$  and from equation (54). If equation (61) is written in matrix form as

$$\mathbf{s}_i = \mathbf{F}_i \mathbf{a}_i, \quad (66)$$

where  $\mathbf{F}_i$  is the  $M \times K_i$  matrix with components  $[F_i]_{mk} = f_{ikm}$ , then  $E(s_i s_i^\dagger)$  is given by

$$E(s_i s_i^\dagger) = \mathbf{F}_i \mathbf{K}_{a_i} \mathbf{F}_i^\dagger. \quad (67)$$

The matrix  $\mathbf{F}_i$  can be compactly expressed, using equation (58), as

$$\mathbf{F}_i = \int_{T_2}^{T_1} \Psi^\dagger(t) \mathbf{F}_i(t) dt, \quad (68)$$

where the  $N \times M$  matrix  $\Psi(t)$  is defined by

$$\Psi(t) = [\Psi_1(t), \Psi_2(t), \dots, \Psi_M(t)]. \quad (69)$$

In summary,  $\mathbf{x}$  is a truncated discrete representation of the process  $\mathbf{x}(t)$  in terms of the complete set of orthogonal vector-valued functions  $\{\Psi_m(t)\}$ :

$$\mathbf{x} = \int_{T_2}^{T_1} \Psi^\dagger(t) \mathbf{x}(t) dt, \quad (70)$$

which is equation (51) in matrix form. Moreover, for all  $i = 0, 1, 2, \dots, I$ ,  $\mathbf{K}_{x_i}$  (defined as the covariance matrix of  $\mathbf{x}$ , i.e.,  $\mathbf{K}_{x_i} = E(\mathbf{x}\mathbf{x}^\dagger | \mathcal{H}_i)$ ) is given by

$$\mathbf{K}_{x_0} = \frac{N_0}{2} \mathbf{I}, \quad (71a)$$

$$\mathbf{K}_{x_i} = \mathbf{F}_i \mathbf{K}_{a_i} \mathbf{F}_i^\dagger + \frac{N_0}{2} \mathbf{I}. \quad (71b)$$

It should be pointed out that, in this case, equation (48) is not the Karhunen-Loève expansion of  $\mathbf{x}(t)$ , since the coefficients  $\{x_m\}$  are not mutually uncorrelated. The vector-valued functions need only form a complete (orthogonal) set; otherwise, they are arbitrary. In general, the Karhunen-Loève expansion of  $\mathbf{x}(t)$  under all  $\mathcal{H}_i$  does not exist for  $I \geq 2$ . The bases for the Karhunen-Loève expansion are the vector-valued eigenfunctions of the covariance matrix  $\mathbf{K}_{x_i}(t_1, t_2)$  that, of course, vary with  $i$ ; i.e., for  $i = 1, 2, \dots, I$ ,

$$\int_{T_2}^{T_1} \mathbf{K}_{x_i}(t, \tau) \mathbf{e}_{im}(\tau) d\tau = \lambda_{im} \mathbf{e}_{im}(t), \quad (72)$$

for  $t$  in  $[T_1, T_2]$  and  $m = 1, 2, \dots, M$ . The vector-valued eigenfunctions  $\{\mathbf{e}_{im}(t)\}$  and the corresponding eigenvalues  $\{\lambda_{im}\}$  are not, in general, independent of  $i$ . Since the *detection* problem is the binary test ( $I = 1$ ), the Karhunen-Loève expansion can be applied (references 3 and 4). However, the classification problem is at least ternary ( $I \geq 2$ ) and coefficients that are mutually uncorrelated under each  $\mathcal{H}_i$  cannot, in general, be found. At any rate, such uncorrelated coefficients are unnecessary for the development herein.

## GAUSSIAN CASE

If  $x(t)$  is a complex Gaussian process then, as is well known, a linear functional like that in equation (70) produces a Gaussian vector  $x$  with mean and covariance matrix given by equations (64) and (71), respectively. Explicitly, the probability density function of  $x$  conditioned on  $\mathcal{H}_i$  is given by,

$$p(x|\mathcal{H}_i) = \frac{1}{\pi^M |K_{x_i}|} \exp[-x^\dagger K_{x_i}^{-1} x], \text{ for } i = 0, 1, 2, \dots, I, \quad (73)$$

where  $|K_{x_i}|$  is the determinant of  $K_{x_i}$  (and the symbol  $x$  is used for the variable in the function  $p(\cdot|\mathcal{H}_i)$  as well for the random (vector) variable).

The likelihood ratio in equation (44) is given by

$$\Lambda_i(x) = \frac{p(x|\mathcal{H}_i)}{p(x|\mathcal{H}_0)}, \quad i = 0, 1, 2, \dots, I,$$

which under the Gaussian assumption becomes

$$\Lambda_i(x) = \frac{|K_{x0}|}{|K_{x_i}|} \exp[-x^\dagger (K_{x_i} - K_{x0}) x] \quad (74)$$

for  $i = 1, 2, \dots, I$ . For the case of continuous sampling, that is, for the process  $x(t)$  for all  $t$  in  $[T_1, T_2]$ , the likelihood ratio is defined as (references 3 and 4)

$$\Lambda_i[x(t)] = \lim_{M \rightarrow \infty} \Lambda_i(x), \quad i = 1, 2, \dots, I. \quad (75)$$

Therefore, it is now necessary to determine

$$\lim_{M \rightarrow \infty} x^\dagger K_{x_i}^{-1} x,$$

and

$$\lim_{M \rightarrow \infty} \frac{|K_{x0}|}{|K_{x_i}|},$$

for  $i = 0, 1, 2, \dots, I$ .

Assume that  $x(t)$  can be expanded in terms of the complete orthonormal (CON) set  $\{\Psi_m(t)\}$  under all  $i = 0, 1, 2, \dots, I$  as in equations (49) and (50). Indeed, this has been shown to be so, at least for the model of  $x(t)$  adopted herein, viz., equation (28). Then, from equation (70),

$$K_{xi} = \int_{T_1}^{T_2} \int_{T_1}^{T_2} \Psi^\dagger(t_1) K_{xi}(t_1, t_2) \Psi(t_2) dt_1 dt_2. \quad (76)$$

Now, express  $K_{xi}^{-1}$  in the following form:

$$K_{xi}^{-1} = \int_{T_1}^{T_2} \int_{T_1}^{T_2} \Psi^\dagger(t_1) Q_{xi}(t_1, t_2) \Psi(t_2) dt_1 dt_2, \quad (77)$$

where the  $(N \times N)$  matrix function  $Q_{xi}(t_1, t_2)$  is to be determined. It can be seen immediately from equation (77) that

$$x^\dagger K_{xi}^{-1} x = \int_{T_1}^{T_2} \int_{T_1}^{T_2} x_M^\dagger(t_1) Q_{xi}(t_1, t_2) x_M(t_2) dt_1 dt_2, \quad (78)$$

since from equation (50), written in vector form,

$$x_M(t) = \Psi(t)x, \quad (79)$$

where the matrix  $\Psi(t)$  is defined by equation (69). Then

$$\lim_{M \rightarrow \infty} x^\dagger K_{xi}^{-1} x = \int_{T_1}^{T_2} \int_{T_1}^{T_2} x^\dagger(t_1) Q_{xi}(t_1, t_2) x(t_2) dt_1 dt_2, \quad (80)$$

for  $i = 0, 1, 2, \dots, I$ .

What remains is to derive an integral equation for  $Q_{xi}(t_1, t_2)$ . As it turns out, the equation is easy to solve for the assumed model of  $x(t)$  under the alternative  $\mathcal{H}_i$  for all  $i = 0, 1, 2, \dots, I$ . Multiply equations (76) and (77) to get

$$I = \int \int \int \int \Psi^\dagger(t_1) K_{x_i}(t_1, t) \Psi(t) \Psi^\dagger(t') Q_{x_i}(t', t_2) \Psi(t_2) dt_1 dt dt' dt_2, \quad (81)$$

where the integrals are each over  $[T_1, T_2]$  and  $I$  is  $M \times M$  in this case. Next recognize that

$$\delta(t_1 - t_2) I = \lim_{M \rightarrow \infty} \sum_{m=1}^M \Psi_m(t_1) \Psi_m^\dagger(t_2) = \lim_{M \rightarrow \infty} \Psi(t_1) \Psi^\dagger(t_2), \quad (82)$$

for any  $t_1, t_2$  in  $[T_1, T_2]$ . Pre-multiplying equation (81) by  $\Psi(t_1)$ , then post-multiplying it by  $\Psi^\dagger(t_2)$ , and taking the limit as  $M \rightarrow \infty$ , yields, invoking equation (82),

$$\delta(t'_1 - t'_2) I = \int \int \int \int \delta(t'_1 - t_1) K_{x_i}(t_1, t) \delta(t - t') Q_{x_i}(t', t_2) \delta(t_2 - t'_2) dt_1 dt dt' dt_2, \quad (83)$$

which reduces to (after dropping primes), for all  $i = 0, 1, 2, \dots, I$ ,

$$\int_{T_1}^{T_2} K_{x_i}(t_1, t) Q_{x_i}(t, t_2) dt = \delta(t_1 - t_2) I, \quad T_1 \leq t_1, t_2 \leq T_2. \quad (84)$$

It can now be seen the  $Q_{x_i}(t_1, t_2)$  is the inverse kernel of  $K_{x_i}(t_1, t_2)$  as defined by the matrix-integral equation (84). Clearly, since  $K_{x_0}(t_1, t_2) = (N_0/2)\delta(t_1 - t_2)I$  for  $t_1, t_2$  in  $[T_1, T_2]$ , then

$$Q_{x_0}(t_1, t_2) = \frac{2}{N_0} \delta(t_1 - t_2) I. \quad (85)$$

Now, for  $i = 1, 2, \dots, I$ , equation (84) becomes, using equation (9) with

$$K_n(t_1, t_2) = (N_0/2)\delta(t_1 - t_2)I,$$

$$\int_{T_1}^{T_2} K_{s_i}(t_1, t) Q_{x_i}(t, t_2) dt + (N_0/2) Q_{x_i}(t_1, t_2) = \delta(t_1 - t_2) I, \quad T_1 \leq t_1, t_2 \leq T_2. \quad (86)$$

Define  $H_i(t_1, t_2)$  by the following equation, for  $i = 1, 2, \dots, I$ :

$$Q_{x_i}(t_1, t_2) = (2/N_0)\delta(t_1 - t_2)I - H_i(t_1, t_2), \quad T \leq t_1, t_2 \leq T_2 \quad (87)$$

Then, equation (86) becomes

$$\int_{T_1}^{T_2} \mathbf{K}_{s_i}(t_1, t) \mathbf{H}_i(t, t_2) dt + \mathbf{H}_i(t_1, t_2) = (2/N_o) \mathbf{K}_{s_i}(t_1, t_2). \quad (88)$$

Equation (88) is the important matrix integral equation that must be solved to gain an explicit expression for the likelihood ratio functional of equation (75). Using equations (80), (85), and (87), it can be seen that  $\Lambda_i[\mathbf{x}(t)]$  becomes

$$\Lambda_i[\mathbf{x}(t)] = \gamma_{0i} \exp \left[ \int_{T_1}^{T_2} \int_{T_1}^{T_2} \mathbf{x}^\dagger(t_1) \mathbf{H}_i(t_1, t_2) \mathbf{x}(t_2) dt_1 dt_2 \right], \quad (89)$$

where

$$\gamma_{0i} = \lim_{M \rightarrow \infty} \frac{|\mathbf{K}_{x_0}|}{|\mathbf{K}_{x_i}|},$$

which can be assumed for now to exist. Equations (88) and (89) define the likelihood ratio that obtains at least for the case when  $\mathbf{x}(t)$  can be expanded in terms of a CON set  $\{\Psi_m(t)\}$  that is independent of  $i = 0, 1, 2, \dots, I$ .

Equation (88) is now solved in closed form for the special case of  $\mathbf{K}_{s_i}(t_1, t_2)$  given by equation (15), viz.,

$$\mathbf{K}_{s_i}(t_1, t_2) = \mathbf{F}_i(t_1) \mathbf{K}_{a_i} \mathbf{F}_i^\dagger(t_2), \quad T_1 \leq t_1, t_2 \leq T, \quad i = 1, 2, \dots, I.$$

It is straightforward to prove that  $\mathbf{H}_i(t_1, t_2)$  is given by

$$\mathbf{H}_i(t_1, t_2) = (2/N_o)^2 \mathbf{F}_i(t_1) \mathbf{H}_i \mathbf{F}_i^\dagger(t_2), \quad (90)$$

where  $\mathbf{H}_i$  is as in equation (35), viz.,

$$\mathbf{H}_i = \mathbf{K}_{a_i} \left[ \mathbf{I} + \frac{2NE_f}{N_o} \Phi_i \mathbf{K}_{a_i} \right]^{-1}.$$



This can be shown by substituting equations (15) and (90) into equation (88) and solving the resultant matrix equation for  $H_i$ . With hindsight, of course, it can be shown that equations (35) and (90) specify the solution by direct substitution into equation (88), given  $K_{s_i}(t_1, t_2)$  as in equation (15). In summary, equation (89), the likelihood ratio functional, becomes

$$\Lambda_i[x(t)] = \gamma_{0i} \exp \left[ \left( \frac{2}{N_o} \right)^2 \int_{T_1}^{T_2} \int_{T_1}^{T_2} x^\dagger(t_1) F_i(t_1) H_i F_i^\dagger(t_2) x(t_2) dt_1 dt_2 \right]. \quad (91)$$

Equation (91) can be written in several alternative forms as follows:

$$\Lambda_i[x(t)] = \gamma_{0i} \exp(y_i^\dagger H_i y_i), \quad (92)$$

where

$$y_i = \frac{2}{N_o} \int_{T_1}^{T_2} F_i^\dagger(t) x(t) dt. \quad (93)$$

Thus,  $y_i$  is the output of a matrix matched filter operation. Another form is

$$\Lambda_i[x(t)] = \gamma_{0i} \exp \left[ \left( \frac{2}{N_o} \right) \int_{T_1}^{T_2} x^\dagger(t) \hat{s}_i(t) dt \right], \quad (94)$$

where

$$\hat{s}_i(t) = F_i(t) \hat{a}_i.$$

Under  $\mathcal{H}_i$ ,  $\hat{a}_i$  is the MVLU (minimum variance, linear, unbiased) estimator of  $a_i$  and is given by equation (37). Thus,  $\hat{s}_i(t)$  is the optimal estimator of  $s_i(t)$  and equation (94) represents the familiar "estimate and match" structure.

### Example: The Ternary Problem

Certain assumptions can now be made regarding the classifier decision statistic as previously given by equation (46) rewritten in terms of the likelihood ratio functional

$$E_i[x(t)] = \sum_{j=0}^I \lambda_{ij} \Lambda_j[x(t)], \quad i = 0, 1, 2, \dots, I. \quad (95)$$

First, let  $c_{ii} = 0$  and  $p_i = p$  where  $0 \leq p \leq 1$  and  $i = 0, 1, 2, \dots, I$ . By this, it is assumed that each hypothesis is equally probable *a priori* and that the costs of correct decisions are zero. Moreover, it is assumed that the cost of deciding  $\mathcal{H}_i$  given  $\mathcal{H}_0$  as true is independent of  $i$ ; i.e., let  $c_{i0} = c_0$ , any non-negative constant, for all  $i = 1, 2, \dots, I$ . With these assumptions, equation (95) becomes for the case of  $I = 2$  the ternary case:

$$E_0[\cdot] = \lambda_{01} \Lambda_1[\cdot] + \lambda_{02} \Lambda_2[\cdot], \quad (96a)$$

$$E_1[\cdot] = \lambda_0 + \lambda_{12} \Lambda_2[\cdot], \quad (96b)$$

$$E_2[\cdot] = \lambda_0 + \lambda_{21} \Lambda_1[\cdot], \quad (96c)$$

where  $\lambda_0 = -c_0 p$ . Using the decision rule prescribed by equations (45) and (46), the following obtains:

$$\text{choose } \mathcal{H}_0 \text{ if } E_1[\cdot] < E_0[\cdot] \text{ and } E_2[\cdot] < E_0[\cdot]; \quad (97a)$$

$$\text{choose } \mathcal{H}_1 \text{ if } E_1[\cdot] > E_0[\cdot] \text{ and } E_1[\cdot] > E_2[\cdot]; \quad (97b)$$

$$\text{choose } \mathcal{H}_2 \text{ if } E_2[\cdot] > E_0[\cdot] \text{ and } E_2[\cdot] > E_1[\cdot]. \quad (97c)$$

In the above comparisons,  $E_i[\cdot] > E_0[\cdot]$ ,  $i = 1, 2, \dots, I$  are tests for *detection* of the corresponding signals  $s_i(t)$ . On the other hand, the comparison  $E_1[\cdot] > E_2[\cdot]$  represents the *classification* of the data  $x(t)$  given that both signals have been correctly detected; it is this test that discriminates between the alternative target classes. By using equations (96b) and (96c), the classifier rule becomes (remembering that  $\lambda_{12}, \lambda_{21} < 0$ ):

$$\text{choose } \mathcal{H}_1 \text{ if } \frac{\Lambda_1[\cdot]}{\Lambda_2[\cdot]} > \frac{\lambda_{12}}{\lambda_{21}}, \text{ otherwise, choose } \mathcal{H}_2. \quad (98)$$

Taking the logarithm, as usual, and using equation (89), inequality (98) becomes:

$$\left(\frac{2}{N_0}\right)^2 \int_{T_1}^{T_2} \int_{T_1}^{T_2} \mathbf{x}^\dagger(t_1) [\mathbf{H}_1(t_1, t_2) - \mathbf{H}_2(t_1, t_2)] \mathbf{x}(t_2) dt_1 dt_2 > l_{12}, \quad (99)$$

where  $l_{12} = \ln (\lambda_{12}\gamma_{02}/\lambda_{21}\gamma_{01})$ . For the special case when  $\mathbf{H}_i(t_1, t_2)$  is as specified by equations (35) and (90), the above becomes, defining  $L_{12}$  as the logarithm of the likelihood ratio,

$$L_{12} = \mathbf{y}^\dagger (\mathbf{H}_1 - \mathbf{H}_2) \mathbf{y} > l_{12}, \quad (100)$$

where the additional assumption is made that  $\mathbf{F}_i(t) = \mathbf{F}(t)$  for  $i = 1, 2$ , and  $\mathbf{y}$  is defined by

$$\mathbf{y} = \frac{2}{N_0} \int_{T_1}^{T_2} \mathbf{F}^\dagger(t) \mathbf{x}(t) dt.$$

This new assumption implies that for each alternative hypothesis (signal class), the time delay set can now be written as  $\{\tau_{nk}\}$ , i.e., independently of the index  $i$ . It is not an overly restrictive assumption, as it turns out. There is flexibility in defining the common test region (as previously defined and illustrated in figure 1), since it is not required for  $\mathbf{K}_{a_i}$  to have full rank.

At this juncture, the distribution of the test statistic  $L_{12}$ , under  $\mathcal{H}_1$  or  $\mathcal{H}_2$ , can be derived, invoking the Gaussian assumption and certain other assumptions. The classifier performance can then be calculated as a function of the critical parameters describing both the system (array, signal design, etc.) and the set of scattering objects. The performance results, viz., operating characteristic curves, will be presented in a later section. But first, properties of the estimator  $\hat{\mathbf{a}}_i$  will be studied further. As shown by equation (94) estimation of  $\hat{\mathbf{a}}_i$  and its error variance matrix (equation 38) will clarify certain of the classifier's properties such as resolution.

## VARIANCE OF $\hat{a}_i$ AND RESOLUTION

The MVLU (under  $\mathcal{H}_i$ ) estimator  $\hat{a}_i$  can be written as, from equations (35), (37), and (93),

$$\hat{a}_i = \mathbf{H}_i \mathbf{y}_i, \quad i=1,2,\dots,I. \quad (101)$$

Note that, given the definition of  $\mathbf{y}_i$ , the following holds:

$$\mathbf{y}_i = \rho_N \Phi_i \mathbf{a}_i + \mathbf{v}_i, \quad (102)$$

where the noise term  $\mathbf{v}_i$  is now given by

$$\mathbf{v}_i = \int_{T_1}^{T_2} \mathbf{F}_i^\dagger(t) \mathbf{n}(t) dt.$$

Equation (102) is characteristic of a class of so-called inverse problems where, in this case, the inverse operator is  $\mathbf{H}_i$ . Observe that when either  $\rho_N \rightarrow \infty$  or  $\mathbf{K}_{a_i}^{-1} \rightarrow 0$ ,  $\mathbf{H}_i$  approaches  $\Phi_i^{-1}$ ; i.e., when *a priori* knowledge is either unnecessary because of high signal-to-noise ratio or is altogether absent,  $\hat{a}_i$  is simply the maximum likelihood estimator\*

$$\hat{a}_i = \frac{1}{\rho_N} \Phi_i^{-1} \mathbf{y}_i. \quad (103)$$

In this case, the inverse operator, while theoretically nonsingular, may be ill-conditioned.

Recall that from the definition of  $\Phi_i$  (equation (19)),

$$[\Phi_i]_{jk} = \frac{1}{N E_f} \sum_{n=1}^N \int_{T_1}^{T_2} f^*(t - \tau_{inj}) f(t - \tau_{ink}) dt, \quad j, k = 1, 2, \dots, K_i, \quad (104a)$$

or, using Parseval's theorem,

---

\*  $\mathbf{K}_{a_i}^{-1} \rightarrow 0$  in the sense that the minimum eigenvalue of  $\mathbf{K}_{a_i} \rightarrow \infty$ .

$$[\Phi_i]_{jk} = \frac{1}{NE_f} \int_{-\infty}^{\infty} |f(\omega)|^2 \sum_{n=1}^n \exp[-i\omega(\tau_{ink} - \tau_{inj})] \frac{d\omega}{2\pi}, \quad (104b)$$

where  $f(\omega)$  denotes the Fourier transform of  $f(t)$ . (To economize on notation, the meaning of  $f(\cdot)$  changes implicitly here depending on whether  $t$  or  $\omega$  appears in the argument.)

From equation (104b) it is clear that, as resolution increases, i.e., as the differences  $\{\tau_{inj} - \tau_{ink}\}$  decrease, the elements of  $\Phi_i$  approach unity, and the matrix becomes ill-conditioned. When the positive definite, Hermitian matrix  $\Phi_i$  is ill-conditioned,  $c(\Phi_i)$ , its so-called condition number, is large;  $c(\Phi_i)$  is defined by  $c(\Phi_i) = \|\Phi_i\| / \|\Phi_i^{-1}\|$ , where  $\|\cdot\|$  is any matrix norm (references 22 and 23). For the 2-norm,  $c(\Phi_i) = \|\Phi_i\|_2 / \|\Phi_i^{-1}\|_2 = \lambda_{ik}/\lambda_{i1}$ , where  $\lambda_{ik}$  and  $\lambda_{i1}$  are the maximum and minimum, respectively, eigenvalues of  $\Phi_i$ . Consider further the maximum likelihood estimator given by equation (103) by examining the sensitivity of  $\hat{a}_i$  to perturbations in both  $\Phi_i$  and  $y_i$ . Let  $\delta\hat{a}_i$  and  $\delta y_i$  be arbitrary perturbations of  $\hat{a}_i$  and  $y_i$ , respectively, and suppose that the perturbation  $\delta\Phi_i$  of  $\Phi_i$  is small enough that  $\alpha < 1$ , where  $\alpha = \|\Phi_i^{-1} \delta\Phi_i\|$ . Then, the following well-known inequality (reference 23) holds (for any vector norm and consistent matrix norm):

$$\frac{\|\delta\hat{a}_i\|}{\|\hat{a}_i\|} \leq (1 - \alpha)^{-1} c(\Phi_i) \left( \frac{\|\delta y_i\|}{\|y_i\|} + \frac{\|\delta\Phi_i\|}{\|\Phi_i\|} \right). \quad (105)$$

Equation (105) reveals the decreased sensitivity of  $\hat{a}_i$  to perturbations both of the data  $y_i$  (viz., measurement noise) and of the matrix  $\Phi_i$  as  $c(\Phi_i)$  decreases. The perturbation of  $\Phi_i$  may represent uncertain variations in the array geometry and test region (e.g., the set  $\{\tau_{ikn}\}$ ). However, the introduction of  $K_{a_i}$  has, in addition to the Bayesian statistical interpretation, the *algebraic* effect of conditioning the inverse  $H_i = [K_{a_i}^{-1} + \rho_N \Phi_i]^{-1}$ , which can be thus written when  $K_{a_i}^{-1}$  exists. Note that for actual computation, the form of  $H_i$  given by equation (35) is used in case  $K_{a_i}$  is singular. These effects will be illustrated in later examples.

It is significant that the estimation problem (or the inverse problem as just discussed) can be physically interpreted as an acoustic imaging process. From equation (38), the variance of the estimation error is, as given previously,

$$\text{var}(\hat{a}_i | \mathcal{H}_i) = H_i \quad i = 1, 2, \dots, I,$$

where equation (35) again is

$$\mathbf{H}_i = \mathbf{K}_{ai} \left[ \mathbf{I} + \rho_N \Phi \mathbf{K}_{ai} \right]^{-1}, \quad i = 1, 2, \dots, I;$$

and  $\Phi_i$ , given by equation (19), can be rewritten, using Parseval's theorem as

$$\Phi_i = \frac{1}{NE_f} \int_{-\infty}^{\infty} \mathbf{F}_i^\dagger(\omega) \mathbf{F}_i(\omega) \frac{d\omega}{2\pi}, \quad (106)$$

where  $\mathbf{F}_i(\omega)$  denotes the Fourier transform of  $\mathbf{F}_i(t)$ . Since the elements of  $\mathbf{F}_i(t)$  are given by  $[\mathbf{F}_i(t)]_{nk} = f(t - \tau_{ink})$ , for  $n = 1, 2, \dots, N$  and  $k = 1, 2, \dots, K_i$ , then equation (106) becomes

$$\Phi_i = \int_{-\infty}^{\infty} |f(\omega)|^2 \mathbf{P}_i^\dagger(\omega) \mathbf{P}_i(\omega) \frac{d\omega}{2\pi}, \quad (107)$$

where  $f(\omega)$ , as before, is the Fourier transform of  $f(t)$  and the  $N \times K_i$  matrix  $\mathbf{P}_i(\omega)$  is defined by

$$[\mathbf{P}_i(\omega)]_{nk} = \exp(-i\omega\tau_{ink}), \quad n = 1, 2, \dots, N; \quad k = 1, 2, \dots, K_i. \quad (108)$$

(Equation (107) is simply equation (104b) with matrix notation.) The matrix  $\mathbf{P}_i(\omega)$  completely defines the geometry of the receiving array of point transducer elements and the geometry of the test region, as previously defined, for the  $i^{\text{th}}$  object. Of course, it is actually the set of time delays  $\{\tau_{ink}\}_1^{K_i}$  that describes the geometry, and these delays are further specified as follows. Referring to figure 2, one can see that for  $i = 1, 2, \dots, I$

$$c\tau_{ink} = \|\mathbf{r}_{ik}\| + \|\mathbf{r}_{ik} - \mathbf{d}_n\|, \quad k = 1, 2, \dots, K_i, \quad n = 1, 2, \dots, N, \quad (109)$$

where  $\mathbf{r}_{ik}$  is the position vector fixing the origin of the  $k^{\text{th}}$  cell of the  $i^{\text{th}}$  object's test region; and  $\mathbf{d}_n$  is the position vector for the  $n^{\text{th}}$  element in the receiving array. A nonrefractive medium with sound speed  $c$  is assumed. A Taylor series expansion in terms of  $\mathbf{d}_n$  about 0 on the right side of equation (109) gives, showing only first order terms,

$$c\tau_{ink} = 2\|\mathbf{r}_{ik}\| - \alpha_{ik}^T \mathbf{d}_n + \dots, \quad (110)$$

where  $\alpha_{ik} = \mathbf{r}_{ik}/\|\mathbf{r}_{ik}\|$ , the unit vector along  $\mathbf{r}_{ik}$ . To neglect higher order terms is to assume that each of the  $K_i$  cells is in the Fraunhofer zone of the array. Note that the second order terms correspond to the Fresnel zone approximation. Here it is assumed that for the array dimensions and distances of interest, the Fraunhofer approximation applies and equation (110) without the higher order terms is taken to be exact. In this case,  $\Phi_i$  becomes

$$[\Phi_i]_{jk} = \frac{1}{NE_f} \int_{-\infty}^{\infty} \left\{ |f(\omega)|^2 \exp[i\omega(\tau_{ij} - \tau_{ik})] \sum_{n=1}^N \exp[-i\omega(\alpha_{ij} - \alpha_{ik})^T \mathbf{d}_n/c] \right\} \frac{d\omega}{2\pi} \quad (111)$$

where  $\tau_{ik} = \frac{2}{c} \|\mathbf{r}_{ik}\|$ ,  $j, k, = 1, 2, \dots, K_i$  and  $n = 1, 2, \dots, N$ .

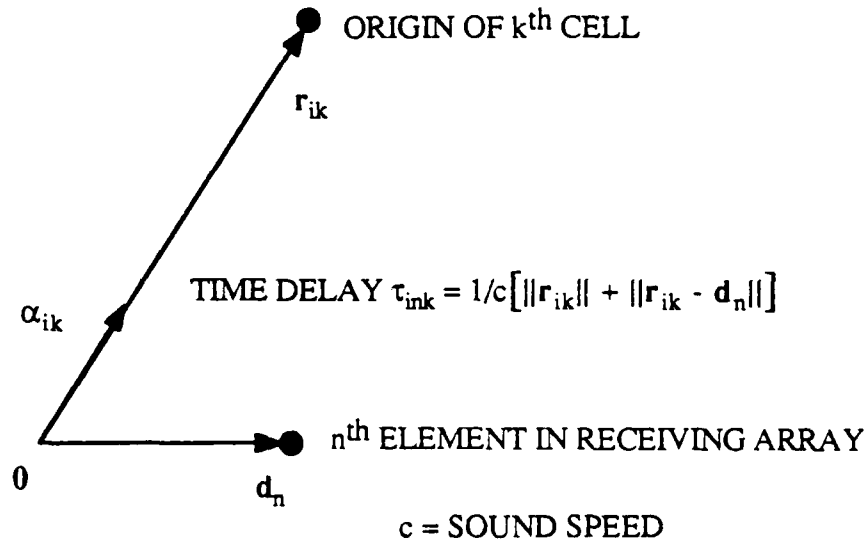


Figure 2. Nomenclature for Nonrefractive Medium

Example: Two-dimensional case and linear array.

Consider the case where  $\alpha_k = [\sin\theta_k, \cos\theta_k, 0]^T$ ,  $k = 1, 2, \dots, K$ . (The subscript  $i$  can be temporarily dropped for this example.) For a uniform linear array,  $\mathbf{d}_n = [(n-1)d, 0, 0]^T$ ,  $n = 1, 2, \dots, N$ , as shown in figure 3.

Further assume in this example that all cells are within a single range annulus, i.e., assume that  $\tau_k = \tau_1$  for all  $k$ . Equation (111) then becomes

$$[\Phi]_{jk} = \overline{NE_f} \int_{-\infty}^{\infty} |f(\omega)|^2 \left[ \frac{\sin\left(\frac{N\omega}{2c} du_{jk}\right)}{\sin\left(\frac{\omega}{2c} du_{jk}\right)} \right] \exp\left[-i \frac{(N-1)\omega}{2c} du_{jk}\right] \frac{d\omega}{2\pi}, \quad (112)$$

where  $u_{jk} = \sin\theta_j - \sin\theta_k$ ,  $j, k = 1, 2, \dots, K$ . Now let the cells be uniform in angular extent; i.e., let  $\theta_k = (k-1)\Delta\theta$ ,  $k = 1, 2, \dots, K$ . The values of  $\Delta\theta$  specify the angular resolution of the processor (equation (103)) with  $\Phi$  given by equation (112)). To proceed with calculations, let  $K_a$  be given by

$$K_a = \sigma_a^2 I. \quad (113)$$

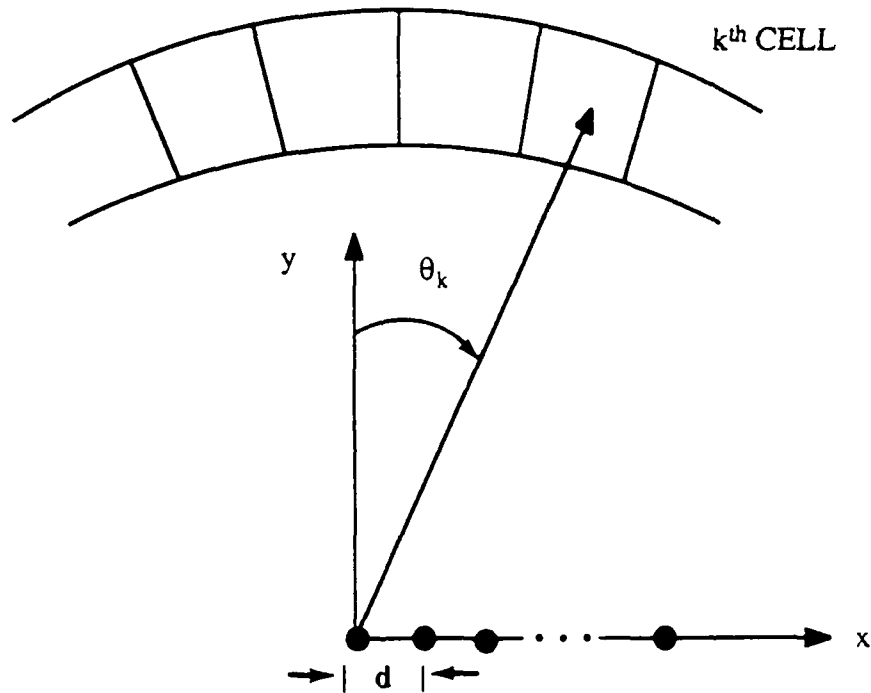


Figure 3. Example: Two-dimensional Case and Linear Array



The example represents the case of scattering strength uniformly distributed, within a range annulus, among  $K$  sectors of angular extent  $\Delta\sigma$ . The variance of  $\hat{\mathbf{a}}$ ,  $\text{var}(\hat{\mathbf{a}})$ , is

$$\text{var}(\hat{\mathbf{a}}) = \mathbf{H} = [\mathbf{K}_a^{-1} + \rho_N \Phi]^{-1}, \quad (114)$$

which, for this example, becomes

$$\text{var}(\hat{\mathbf{a}}) = \sigma_a^2 [\mathbf{I} + \rho_N \sigma_a^2 \Phi]^{-1}. \quad (115)$$

The quantity of interest here is termed the variance ratio, the trace of the error variance matrix normalized by the trace of  $\mathbf{K}_a$ , and is given by, for this example,

$$\frac{\text{tr}[\text{var}(\hat{\mathbf{a}})]}{K\sigma_a^2} = \frac{1}{K} \text{tr} [\mathbf{I} + \rho_N \sigma_a^2 \Phi]^{-1}, \quad (116)$$

where  $\Phi$  is given by equation (112). The variance ratio provides a means of calculating a measure of the variance reduction achieved by the processor (i.e.,  $\hat{\mathbf{a}}$ , the estimator of  $\mathbf{a}$ ) as a function of important parameters such as  $\rho_N \sigma_a^2$  (received signal-to-noise ratio) and the resolution parameter  $\Delta\theta$ . For the example, the variance ratio is calculated (see figure 4) for  $N = 8$ ,  $\omega_0 d/c = \pi$ ,  $K = 2, 3, 8$  (angle cells each of width  $\Delta\theta$ ). It is assumed that  $|f(\omega)|^2$  is narrowband at center frequency  $\omega_0$  and is approximated by  $E_f \delta(\omega - \omega_0)$  in equation (113). The results of these calculations are shown in figure 5. Note that for a uniform linear array, the beamwidth BW, as measured between the first nulls, is given by, for center wavelength  $\lambda_0$ ,

$$\text{BW} = 2 \sin^{-1} \left( \frac{\lambda_0}{Nd} \right), \quad (117)$$

and is, for this example  $\text{BW} = 29^\circ$ . The curves in figure 5 clearly illustrate the tradeoff between resolution and error variance; as the processor attempts greater angular resolution, i.e., as  $\Delta\theta$  decreases, the error variance increases. The tradeoff is more pronounced as the received signal-to-noise ratio  $\rho_N \sigma_a^2$  increases and as the number of cells increases. To precisely specify angular resolution in terms of  $\Delta\theta$ , the variance ratio  $\text{tr}[\text{var}(\hat{\mathbf{a}})]/\text{tr}[\mathbf{K}_a]$  and the number of cells must be first specified. Corresponding values of  $\Delta\theta$  as a function of  $\rho_N \sigma_a^2$  or, more practically, increments of

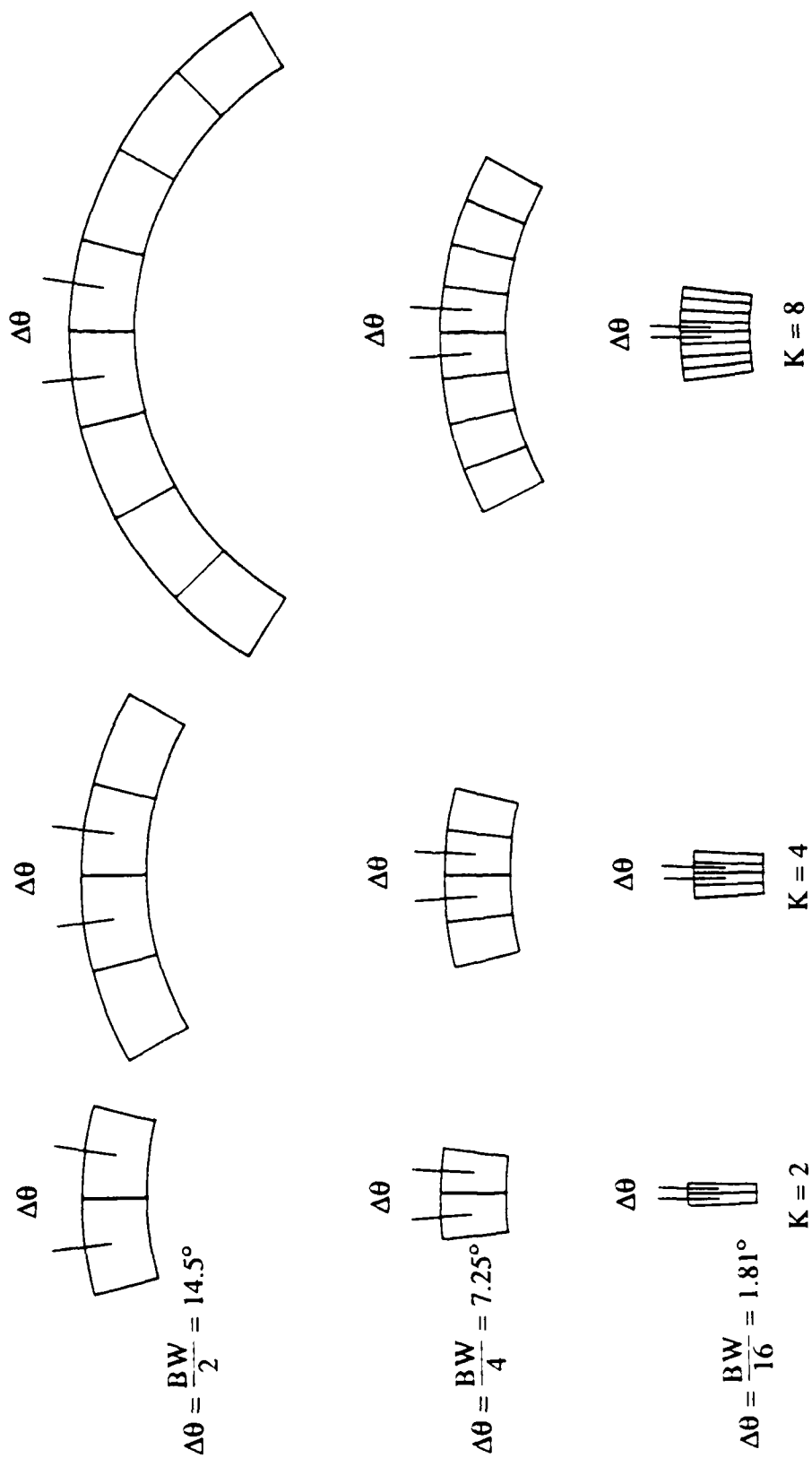


Figure 4. Test Regions for Example: Two-dimensional Linear Array

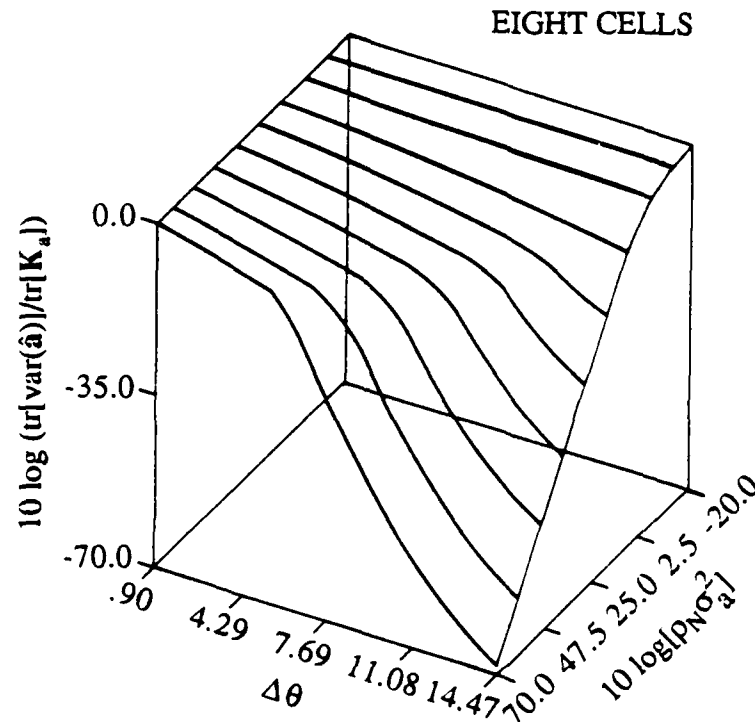
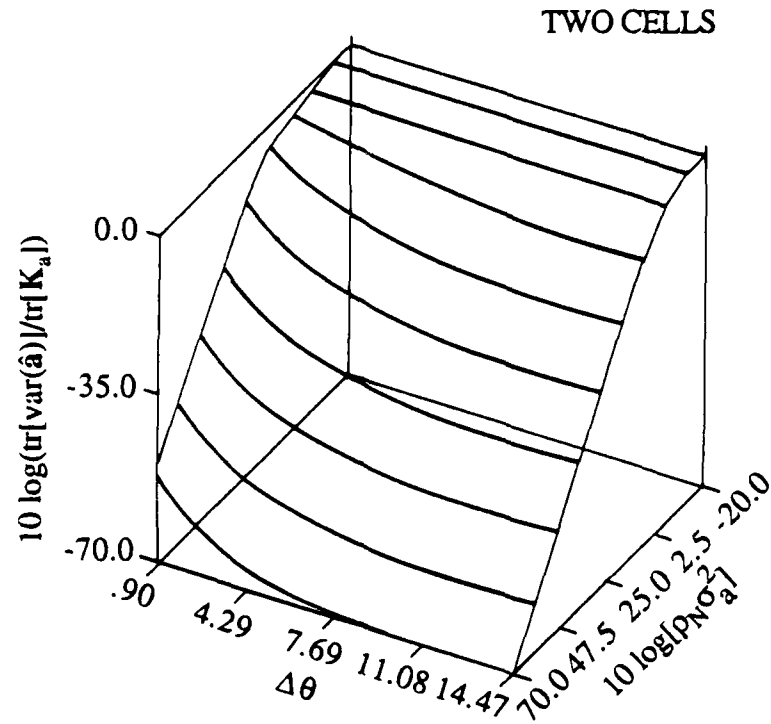


Figure 5. Variance Ratio  $\text{tr}[\text{var}(\hat{a})]/\text{tr}[K_a]$  vs Angle Cell Width  $\Delta\theta$  and Received Signal-to-Noise Ratio  $\rho_N \sigma_a^2$  (Two and Eight Angle Cells)

$\rho_N \sigma_a^2$  can be determined. For example, referring to figures 6 and 7, if a variance ratio of  $10^{-2}$  (-20 dB) is specified, then for  $K = 2$  cells a change in resolution from  $\Delta\theta = BW/2$  ( $= 14.5^\circ$ ) to  $\Delta\theta = BW/8$  ( $= 3.62^\circ$ ), a fourfold improvement, requires an increase in  $\rho_N \sigma_a^2$  of about 6 dB. However, to achieve this same improvement, but for  $K = 8$  cells,  $\rho_N \sigma_a^2$  must be increased by about 60 dB. This example reveals the strong dependence of resolution on the *number* of cells to be resolved. Note that in both cases,  $K = 2$  and  $K = 8$ , the curve for  $\Delta\theta = BW/2$  ( $= 14.5^\circ$ ) nearly coincides with the curve corresponding to  $\Phi = I$ , which represents so-called ideal resolution and is independent of  $K$ ; in this case, the variance ratio is given simply by

$$\frac{\text{tr}[\text{var}(\hat{a})]}{\text{tr}[K_a]} = \frac{1}{1 + \rho_N \sigma_a^2} \quad (118)$$

Further insight into the resolving power of the combination of a particular array and waveform is gained by an eigenanalysis of the correlation matrix  $\Phi$ . In this example,  $\Phi$  is a  $K \times K$  matrix with elements given by equation (112). In general,  $\Phi$  is Hermitian and positive definite, and thus has spectral decomposition given by

$$\Phi = \sum_{k=1}^K \lambda_k \mathbf{e}_k \mathbf{e}_k^\dagger, \quad (119)$$

where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K$  are the ordered, real positive eigenvalues of  $\Phi$ , and  $\{\mathbf{e}_k\}$  are the corresponding eigenvectors for  $n=1,2,\dots,K$ .

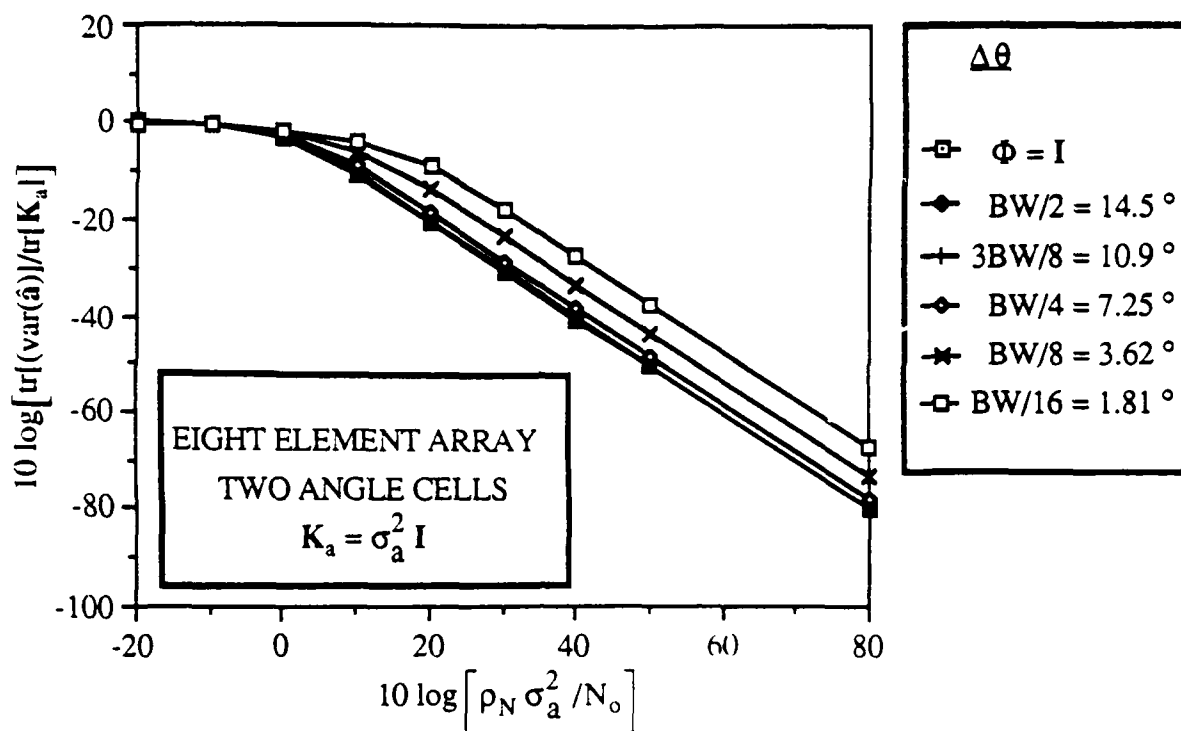


Figure 6. Variance Ratio  $\text{tr}[\text{var}(\hat{a})]/\text{tr}[K_a]$  vs Angle Cell Width  $\Delta\theta$  and Received Signal-to-Noise Ratio  $\rho_N \sigma_a^2$  (Two Angle Cells)

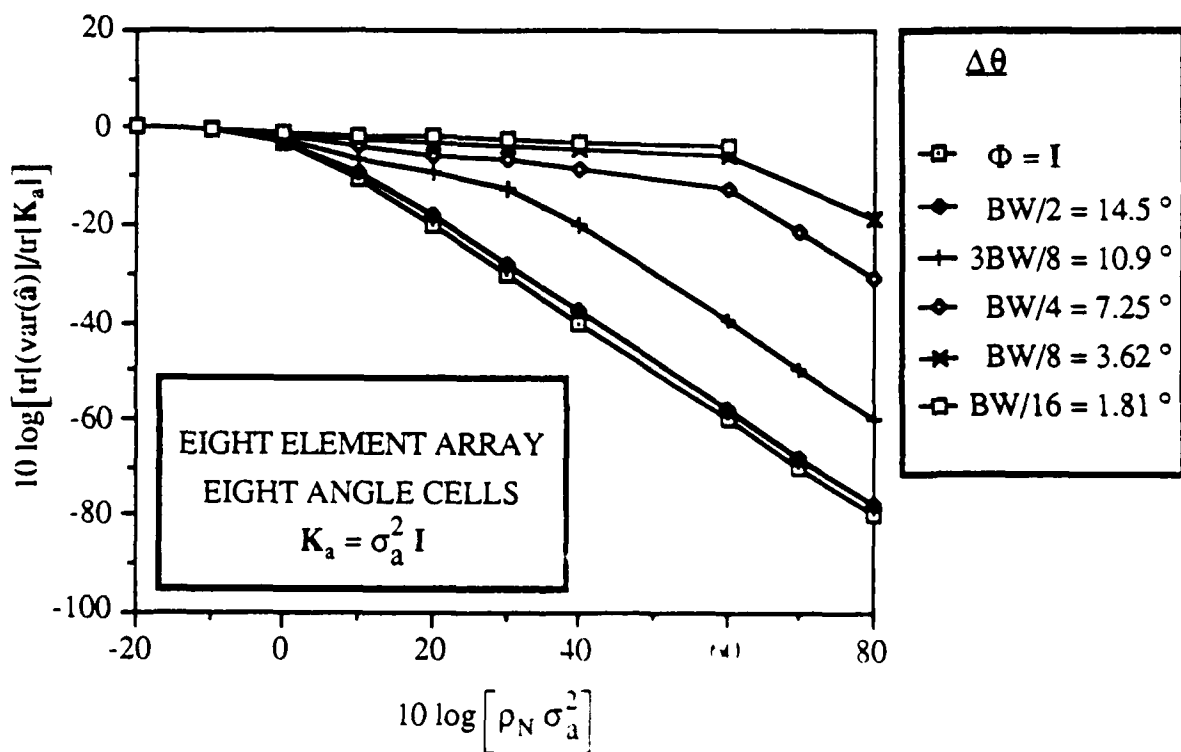


Figure 7. Variance Ratio  $\text{tr}[\text{var}(\hat{a})]/\text{tr}[K_a]$  vs Angle Cell Width  $\Delta\theta$  and Received Signal-to-Noise Ratio  $\rho_N \sigma_a^2$  (Eight Angle Cells)

Figure 8 shows  $\lambda_k/\lambda_1$  versus  $k$  and  $\Delta\theta$  for the narrowband case ( $|f(\omega)^2| = E_f\delta(\omega - \omega_0)$ ). As  $\Delta\theta$  decreases, the condition number  $C[\Phi] = \lambda_1/\lambda_K$  increases rapidly and  $\Phi$  becomes sensitive to noise (and other perturbations as shown in reference 22). However, as shown previously in this example, the introduction of *a priori* information in the form of a  $K_a$  can, with sufficient received signal-to-noise ratio  $\rho_N\sigma_a^2$ , reduce the variance of  $\hat{a}$ , thereby compensating for an ill-conditioned  $\Phi$ .

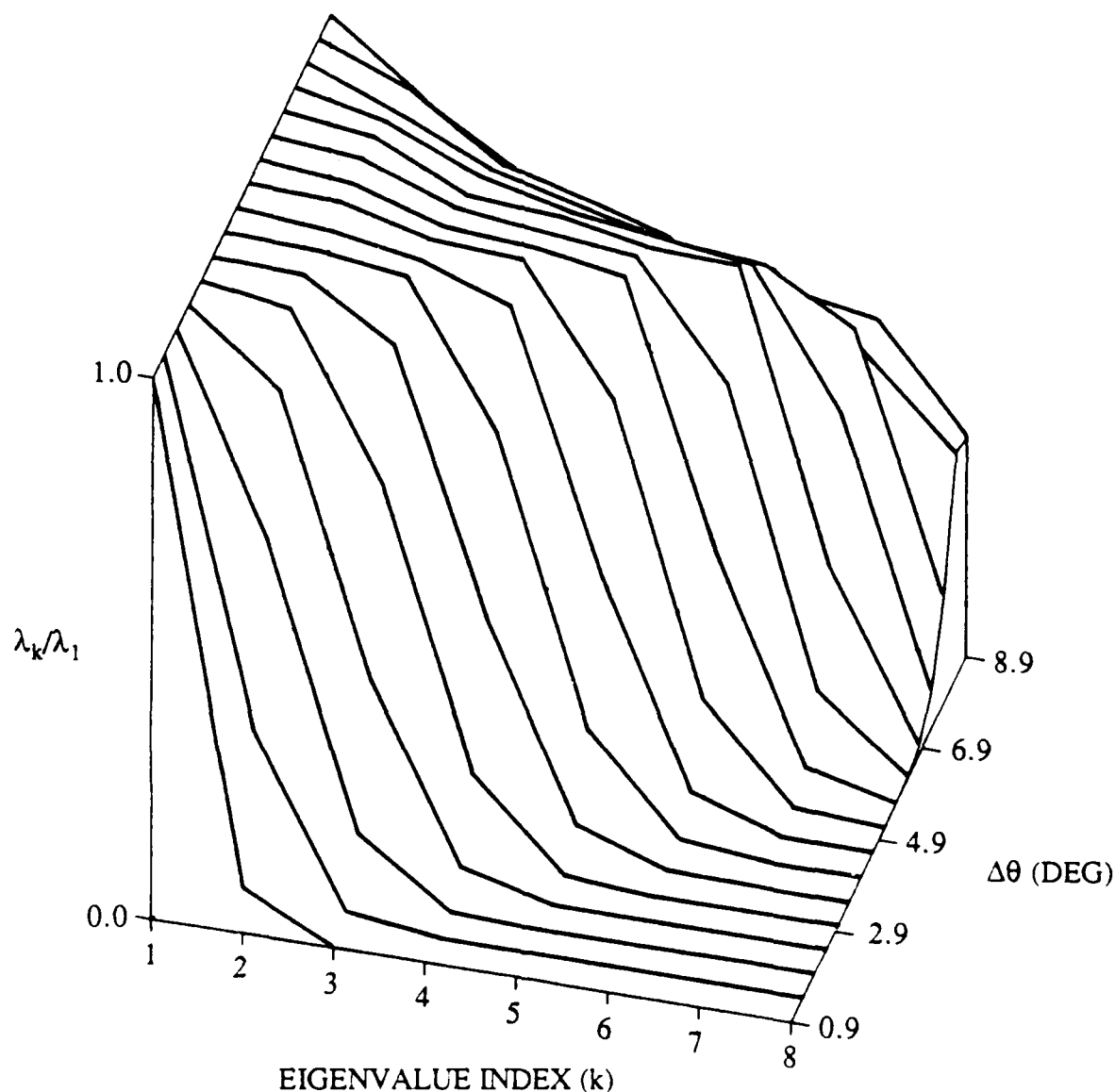


Figure 8. Eigenvalue Spread for Narrowband  $\Phi$

Up to now, the waveform in this example has been assumed to be narrowband and approximated by  $E_f \delta(\omega - \omega_0)$ . To examine the effect of bandwidth, assume that  $|f(\omega)|^2$  is given by

$$|f(\omega)|^2 = \begin{cases} (2\pi B)^{-1}, & |\omega - \omega_0| \leq \pi B, \\ 0, & |\omega - \omega_0| > \pi B, \end{cases} \quad (120)$$

where  $\omega_0$  and  $B$  are the center frequency and bandwidth, respectively. Figures 9 through 11 show  $(\lambda_k/\lambda_1)$  versus  $k$  and  $B/f_0$  for various values of  $\Delta\theta$  ( $BW/8, BW/4, 3BW/8$ ). In this case,  $K = 16$  to determine if increased bandwidth can (in this example) significantly increase values of  $\lambda_k$  for  $k > N = 8$ . Note that the rank of  $\Phi$  is minimal ( $N, K$ ) for the narrowband approximation (actually single frequency). This is clear from equation (107), since  $P(\omega)$  is  $N \times K$ . For the case at hand, i.e., for  $N = 8, K = 16$  increased bandwidth does not significantly amplify  $\lambda_k$  for  $k > 8$ , except for values of  $B/f_0$  impractically large for the kinds of sonar systems under consideration. Of course, this conclusion applies to this example and is qualified accordingly. Further examination of equation (107) should produce generalizations regarding the conditionedness of  $\Phi$  as it depends on array and waveform design.

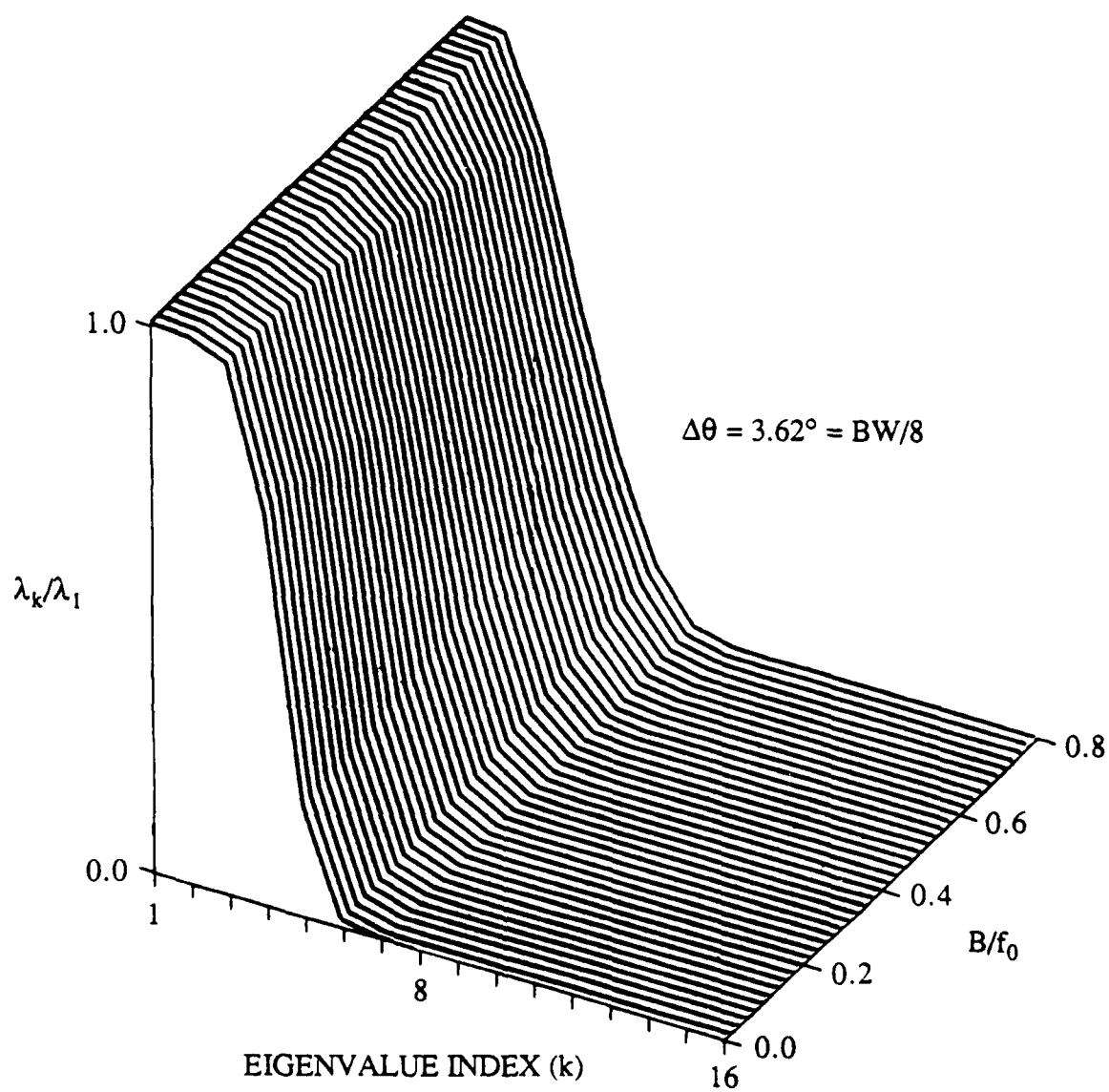


Figure 9. Eigenvalue Spread for Wideband  $\Phi$  ( $\Delta\theta = 3.62^\circ = BW/8$ )



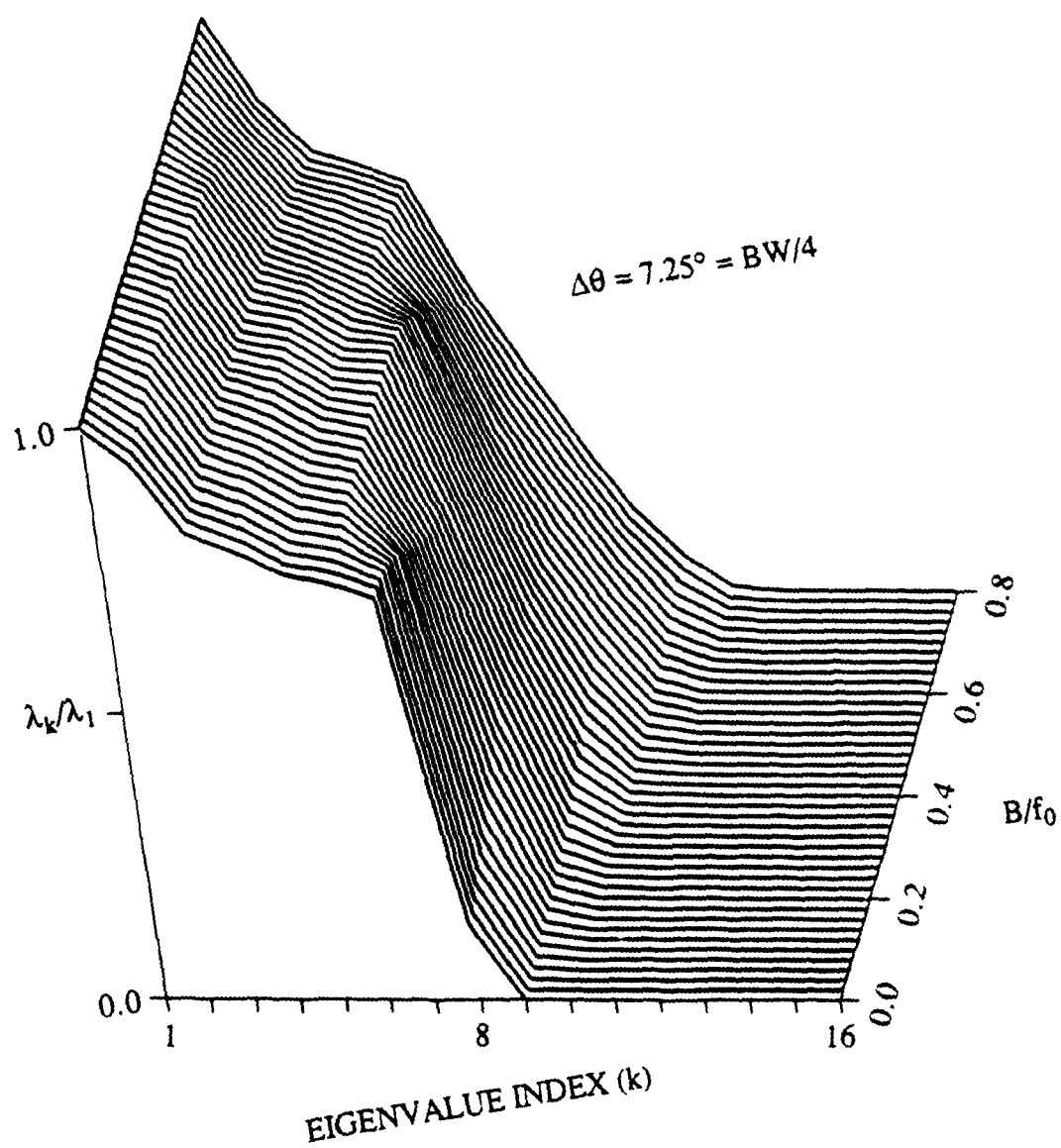


Figure 10. Eigenvalue Spread for Wideband  $\Phi$  ( $\Delta\theta = 7.25^\circ = BW/4$ )

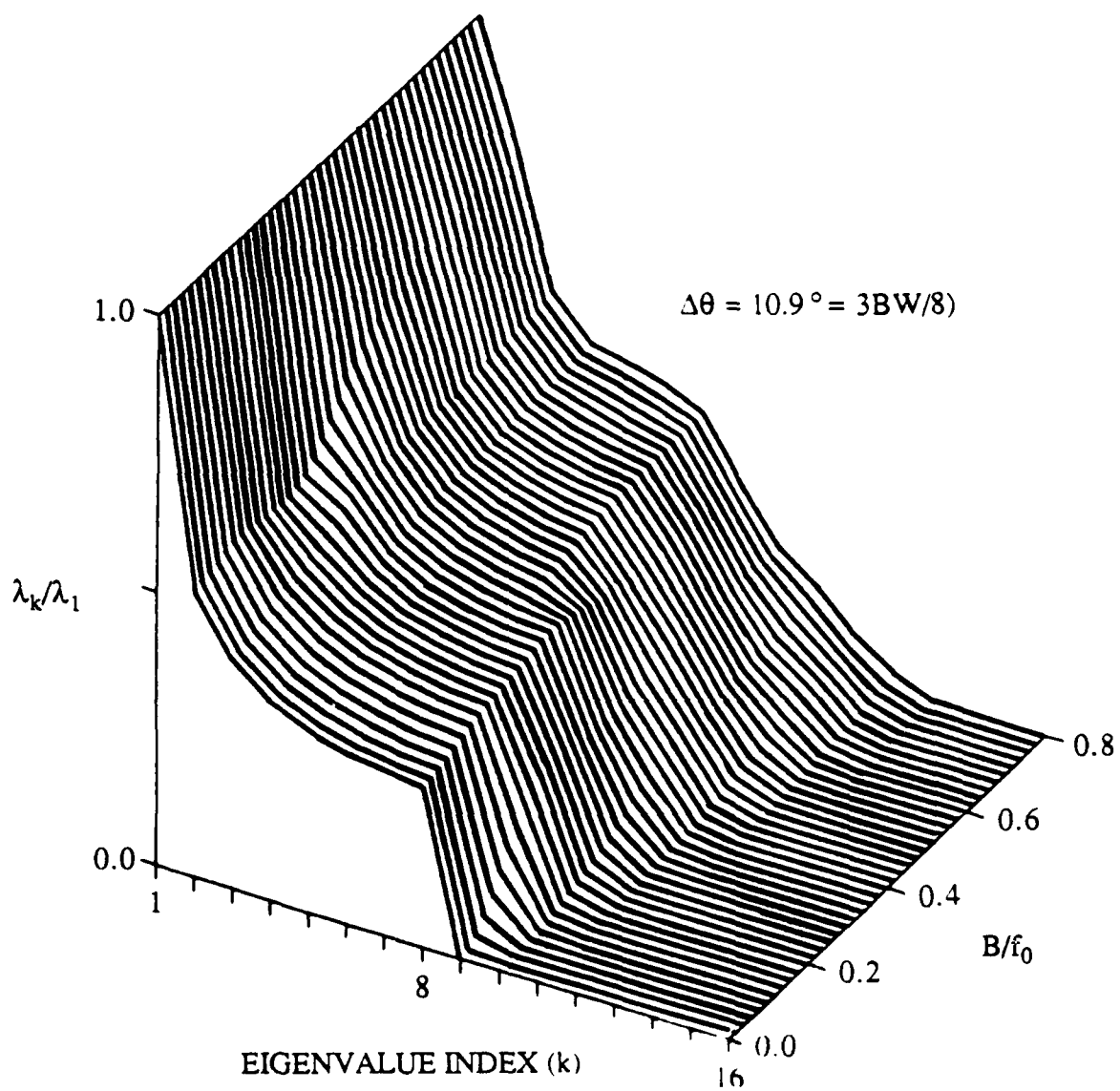


Figure 11. Eigenvalue Spread for Wideband  $\Phi$  ( $\Delta\theta = 10.9^\circ = 3BW/8$ )

## CLASSIFIER PERFORMANCE

This section outlines the performance of the optimum processor for the binary classification problem. It briefly covers some of the components of performance prediction, and then presents selected results. Both the details of classification performance prediction and a large number of performance examples are developed in appendix A.

A brief review of the conditions of the problem is given first. The test region is assumed to be identical under each hypothesis; each scattering coefficient covariance matrix is assumed to have full rank, and the noise is assumed to be both temporally and spatially white. For the binary problem there are two hypotheses:  $\mathcal{H}_1$  and  $\mathcal{H}_2$ . Under these conditions, the sufficient statistic for deciding whether the received signal was from object 1 or 2 was given in equation (100) (a number of equations will be repeated here for continuity),

$$L = \mathbf{y}^\dagger (\mathbf{H}_1 - \mathbf{H}_2) \mathbf{y}$$

where  $\mathbf{y}$  is given in equation (93),

$$\mathbf{y} = \frac{2}{N_0} \int_0^T \mathbf{F}^\dagger(t) \mathbf{x}(t) dt$$

and, from equation (35),

$$\mathbf{H}_i = \mathbf{K}_{a_i} [\mathbf{I} + \rho_N \Phi \mathbf{K}_{a_i}]^{-1} = [\mathbf{K}_{a_i}^{-1} + \rho_N \Phi]^{-1}, \quad i = 1, 2.$$

Classification performance is a function of the distribution of the scalar random variable  $L$ , under  $\mathbf{H}_1$  or  $\mathbf{H}_2$ . If  $p_{ij}$  is defined to be the probability of selecting object  $i$  when object  $j$  occurred, then

$$p_{11}(\eta) = \int_{\eta}^{\infty} p(L|\mathcal{H}_1) dL, \tag{121a}$$

$$p_{12}(\eta) = \int_{\eta}^{\infty} p(L|\mathcal{H}_2)dL, \quad (121b)$$

where  $p(L|\mathcal{H}_i)$  is the probability density function of  $L$  conditioned on the  $i^{\text{th}}$  hypothesis ( $i = 1, 2$ ), and  $\eta$  is the decision threshold.  $P_{11}$  is a measurement of correct classification, while  $P_{12}$  yields a measure of incorrect classification.

It is shown in appendix B that  $L$  is a complex quadratic form in the Gaussian vector  $y$  (note that the matrix  $H_1 - H_2$  is Hermitian but not necessarily positive definite). The characteristic function of quadratic forms in complex Gaussian vectors is well known and for the current problem it can be shown to be (reference 23):

$$M_i(j\omega) \equiv E[\exp(j\omega L)|\mathcal{H}_i] = [\det(C_i(j\omega))]^{-1}, \quad i = 1, 2, \quad (122a)$$

where

$$C_i(j\omega) = I - j\omega K_{y_i}(H_1 - H_2), \quad i = 1, 2, \quad (122b)$$

and  $\det(\cdot)$  is the determinant.

The probability density functions of interest can then be written as

$$p(L|\mathcal{H}_i) = \frac{1}{2\pi} \int_{-\infty}^{\infty} M_i(j\omega) \exp(-j\omega L) d\omega, \quad i = 1, 2. \quad (123)$$

In general, considering the complexity of the matrices  $H_i$  ( $i = 1, 2$ ) any performance evaluation based on equations (122) would have to be made on a per case basis by numerical means and would not necessarily yield any insight into the classification performance of the processor in general. If the following simplifying assumptions are made, some insight into the processor performance may be gained. First, assume that the waveform and the array are capable of perfect resolution in range and angle (for the particular test region assumed). That is, assume

$$\Phi = I. \quad (124)$$

Also assume that, under each hypothesis, the  $K$  scatterers in the test volume are statistically independent so the  $K_{ai}$  now represents a discrete version of the conventional scattering function (reference 2). Thus,  $K_{ai}$  is a diagonal matrix ( $i = 1, 2$ ).

Given these assumptions, it can be shown that, under each hypothesis,  $L$  consists of the sum of  $K$  independent random variables:

$$L = \sum_{k=1}^K l_k,$$

$$\text{where } l_k = \left( \frac{\sigma_{k1}^2}{1 + \sigma_{k1}^2 \rho_N} - \frac{\sigma_{k2}^2}{1 + \sigma_{k1}^2 \rho_N} \right) y_k^2. \quad (125)$$

Here  $y_k$  is the  $k^{\text{th}}$  element of the vector  $y$ ,  $\rho_N$  is the input signal-to-noise ratio  $\rho_N = 2NE_f/N_o$ , and  $\sigma_{ki}^2$  is the  $k^{\text{th}}$  diagonal element of  $K_{ai}$  (scattering strength of the  $k^{\text{th}}$  cell under the  $i^{\text{th}}$  hypothesis). Thus, under each of the two hypotheses,  $L$  is equal to the sum and/or difference (note that the bracketed term can be positive or negative) of independent, not necessarily identically distributed, exponential random variables. Its density functions  $p(L|\mathcal{H}_i)$ ,  $i = 1, 2$ , can be found from equation (123) via residue theory. These can be used to solve for  $P_{ij}$  ( $i, j = 1, 2$ ) of equations (121).

Before addressing a specific set of target descriptions, some additional points should be made. The test volume is assumed to be identical for each hypothesis, and comprises  $K$  cells with an independent specular scatterer in each cell. If under a specific hypothesis no target is present, and the scattered return is due to reverberation alone, all cells in the test volume will possess uniform scattering strength. If, in addition to reverberation, a target is present under a specific hypothesis, the uniform scattering strength of a number of the cells ( $\leq K$ ) in the test volume may be replaced with new scattering strengths; the number of cells, their location, and the magnitude of their scattering strengths will vary according to the target. Thus, it is assumed that each individual scatterer falls into one of two categories, target-like or reverberation, and the difference between hypotheses  $\mathcal{H}_1$  and  $\mathcal{H}_2$  is due to the number, location, and scattering strength of the target-like scatterers. Thus, although their number may be different, it is assumed in these examples that reverberation scatterers are of equal strength under both hypotheses.

Because perfect resolution has been assumed, it is possible to index the  $K$  cells arbitrarily, but identically, under both  $\mathcal{H}_1$  and  $\mathcal{H}_2$ . Thus, the following convention will be adopted: from the  $K$  test cells, find the total  $J$ , such that  $J \leq K$  and

$$\sigma_{j1}^2 \neq \sigma_{j2}^2, \quad j = 1, 2, \dots, J. \quad (126)$$

The remaining  $K - J$  cells of equal scattering strength do not contribute to equation (125). The specific ordering of the  $J$  cells with unequal scattering strengths can be arbitrary as mentioned before. In light of this, the intersection of reverberation cells under  $\mathcal{H}_1$  and  $\mathcal{H}_2$  does not enter into the problem; this is also the case for the intersection of any target-like cells of equal strength.

As an example, assume that each target is comprised of scatterers of constant strength. Under hypothesis 1, the target consists of  $M$  cells, each with a receive signal-to-noise ratio ( $\sigma_{k1}^2 \rho_N$ ,  $k = 1, \dots, M$ ) of  $\beta$ . Under hypothesis 2, the target comprises  $J > M$  cells, the same  $M$  cells as under hypothesis 1, and  $J - M$  additional cells, each with a receive signal-to-noise ratio ( $\sigma_{k2}^2 \rho_N$ ,  $k = 1, \dots, J$ ) of  $\Psi$  (see figure 12 for an example of the test volume geometry). In addition, it is assumed that the total target scattering strength under each hypothesis is equal ( $M\beta = J\Psi$ ). Due to the unequal number of target scatterers, a number  $J - M$  of reverberation scatterers, each with a receive signal-to-noise ratio equal to  $\gamma$ , must be accounted for.

In figure 13 performance curves have been plotted for some representative values of  $\beta$ ,  $\psi$ ,  $\gamma$ ,  $M$ , and  $J$ . Here  $M$  and  $\beta$  are held constant, and  $J$  is increased, thus  $\psi$  is decreased. Reverberation is set equal to a constant value in each curve ( $\gamma = \psi/4$ ). The results indicate that, for a given  $P_{12}$  (probability of incorrect classification), there is a particular set of targets, characterized by some  $M'$ ,  $\beta'$ , and  $J'$ ,  $\psi'$ , for which the spreading of the fixed total receive energies,  $M'\beta'$ , and  $J'\psi'$ , make them the most easily distinguished among all of the targets with constant total signal-to-noise ratios. This result, the spreading of the total fixed receive energy over a particular number of independent channels (cells) for maximum performance, is an example of the well-known "diversity" phenomenon (reference 2).

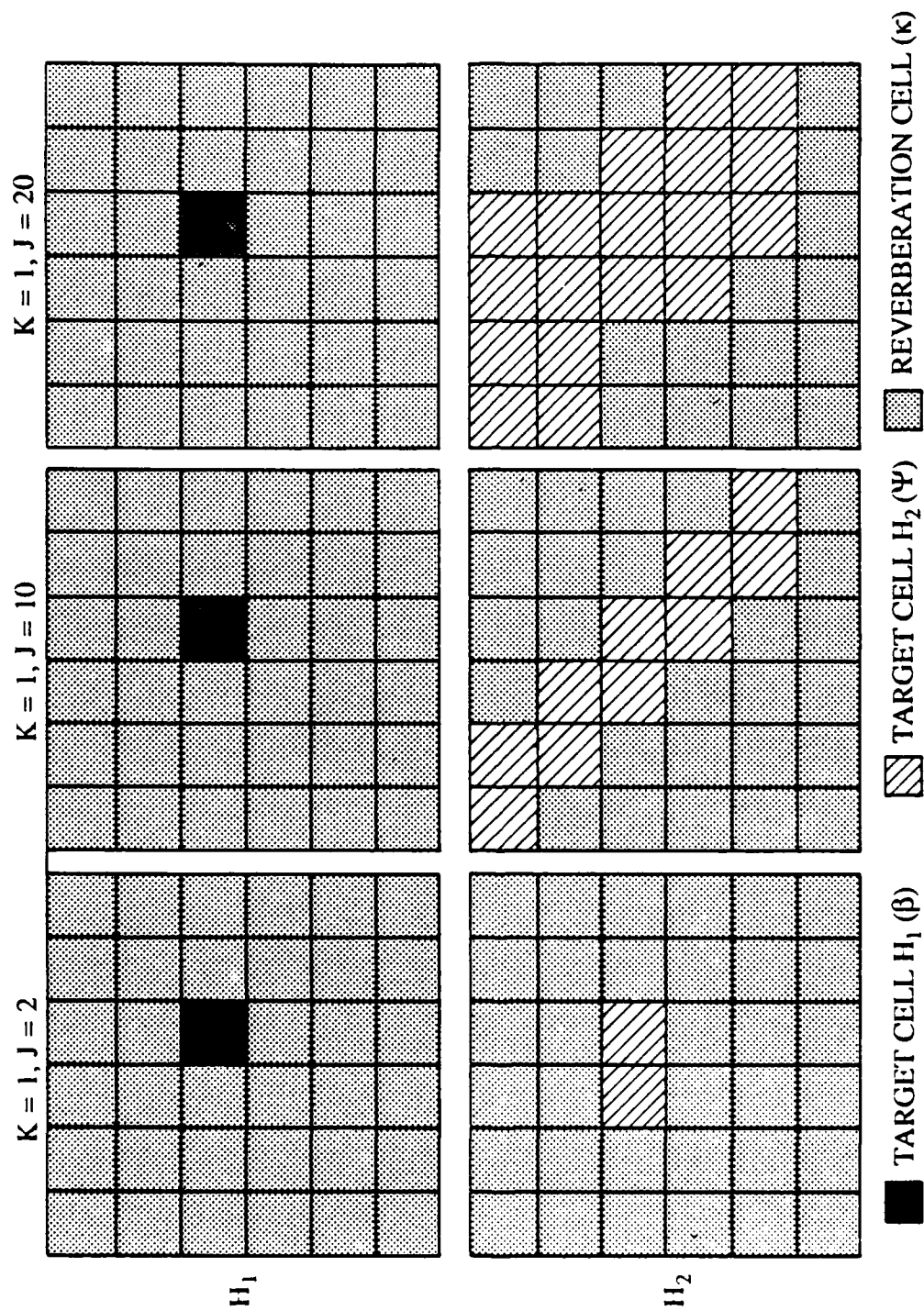


Figure 12. Example of a Two-dimensional Test Volume Geometry;  $K = 36$  Test Cells;

$M = 1$  Target Cell Under  $H_1$ ;  $J = 2, 10, 20$  Target Cells Under  $H_2$

(Note: Target cell locations under  $H_1$  are a subset of those under  $H_2$ .)

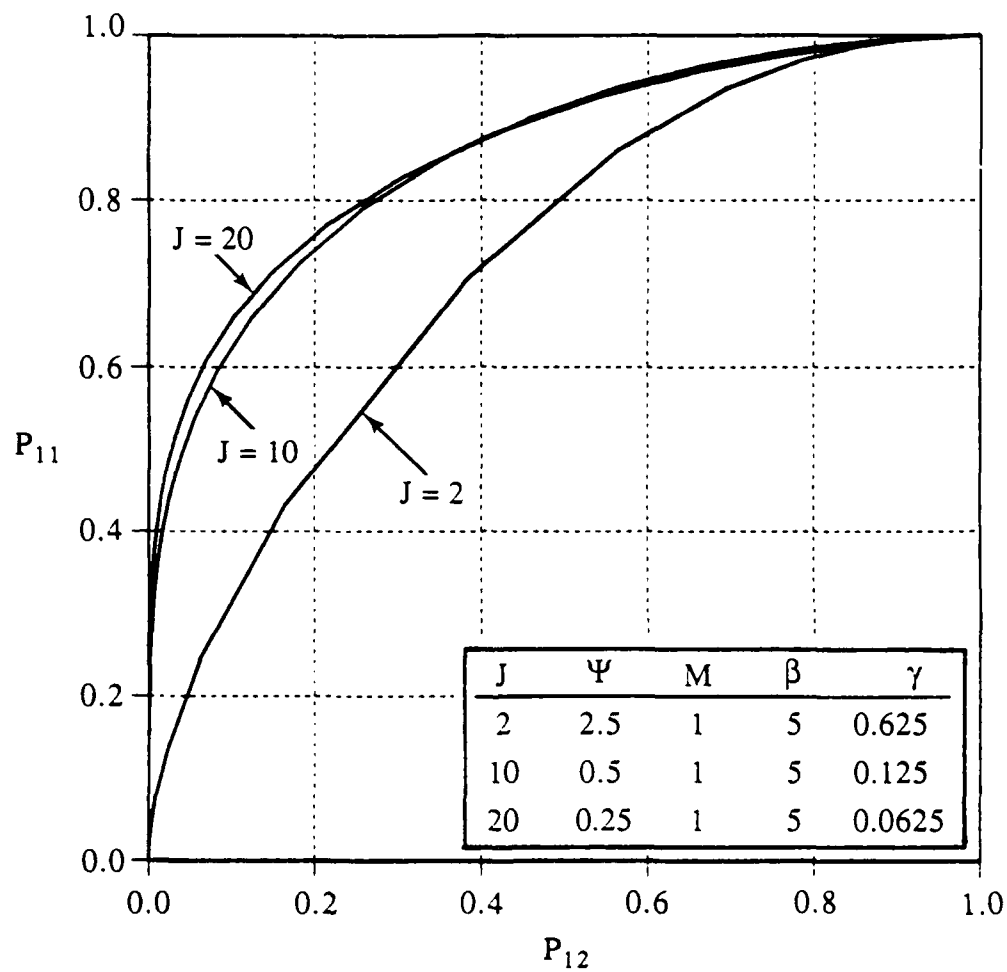


Figure 13. Receiver Operating Characteristic (ROC) Curve;  
 $MB = J\Psi = 5$ ;  $m = 1$ ;  $\gamma = \Psi/4$ ; vary  $J = 2, 10, 20$



## EXTENSIONS

Important extensions of this work will be considered for future research. Several are discussed below.

### MOVING OBJECTS

Throughout this report, it has been assumed that no relative motion exists between the array and the scattering objects. Consider now the case of a rigid body moving with constant velocity; that is, if  $V_{ik}(t)$  is the velocity of the  $k^{\text{th}}$  cell corresponding to the  $i^{\text{th}}$  object, then for a rigid body

$$V_{ik}(t) = V_i(t), \quad k = 1, 2, \dots, K_i, \quad (127)$$

and for uniform motion

$$V_i(t) = V_i, \quad T_1 \leq t \leq T_2. \quad (128)$$

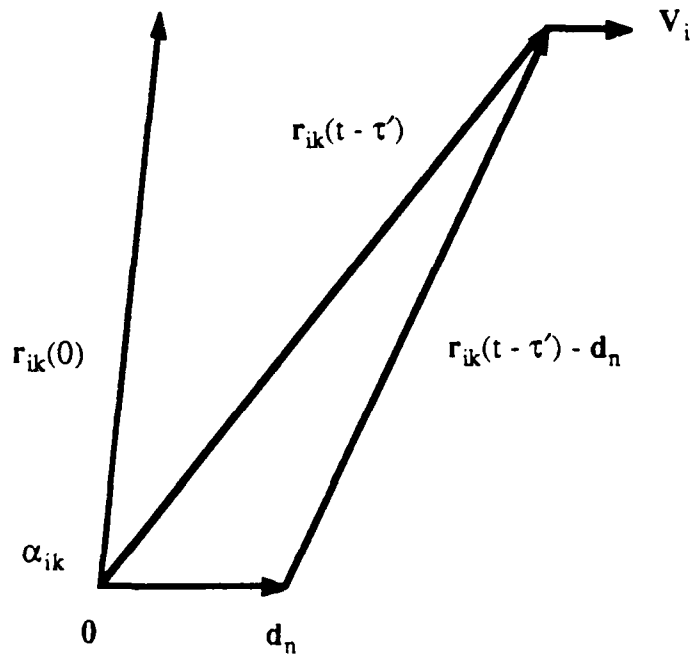


Figure 14. Nomenclature for Object Motion

As can be seen in figure 14,  $V_i$  is relative to a coordinate system with its origin fixed at the receiving array's reference. Moreover, the array elements are assumed to be stationary within this coordinate system. The signal scattered from  $i^{\text{th}}$  object and received at the  $n^{\text{th}}$  array element is now given by

$$s_{in}(t) = \sum_{k=1}^{K_i} a_{ik} f[t - \tau_{ink}(t)], \quad T_1 \leq t \leq T_2, \quad (129)$$

where the  $\tau_{ink}(t)$  is defined as in equation (11) except that it is now a function of time.

An expression for  $\tau_{ink}(t)$  is developed here for completeness. First, for notational simplicity suppress for now the subscripts and time dependence of  $\tau_{ink}(t)$ , denoting it simply by  $\tau$ . Then,

$$\tau = \tau' + \tau''. \quad (130)$$

where the delay  $\tau'$  is the return delay (from scatterer to array element) and is the implicit solution of the equation

$$c\tau' = \|\mathbf{r}_k(t - \tau') - \mathbf{d}_n\|; \quad (131)$$

and the delay  $\tau''$  is the delay from the array reference 0 to the scattering cell and is given by

$$c\tau'' = \|\mathbf{r}_k(t - \tau')\|, \quad k = 1, 2, \dots, K; \quad n = 1, 2, \dots, N. \quad (132)$$

(Note that the subscript  $i$  has been suspended for now.) An approximate solution to the above obtains as follows. Let  $\mathbf{V} = \mathbf{V}_{rk} + \mathbf{V}_{ck}$  where  $\mathbf{V}_{rk} = (\mathbf{V} \cdot \alpha_k)\alpha_k$  and  $\alpha_k$  is, as before, the unit vector  $\mathbf{r}_k / \|\mathbf{r}_k\|$  and  $\mathbf{r}_k = \mathbf{r}_k(0)$ . Thus,  $\mathbf{V}_{rk}$  and  $\mathbf{V}_{ck}$  are the (orthogonal) down-range and cross-range, components of  $\mathbf{V}_k$  respectively.

Since

$$\mathbf{r}_k(t) = \mathbf{V}t + \mathbf{r}_k, \quad T_1 \leq t \leq T_2, \quad (133)$$

---

\* Amplitude scaling by  $(1 - \frac{d\tau_{ink}(t)}{dt})^{-1}$  to conserve the waveform's energy is neglected.

then

$$\mathbf{r}_k(t) = (r_k + V_{rk}t)\alpha_k + V_{ck}t \quad (134)$$

where  $r_k = \|\mathbf{r}_k\|$  and  $V_{rk} = \|\mathbf{V}_{rk}\|$ .

Defining  $\mathbf{d}_{cnk}$  by  $\mathbf{d}_{cnk} = \mathbf{d}_n - (\alpha_k \cdot \mathbf{d}_n)\alpha_k$  and substituting equation (134) into equation (131) yields

$$c\tau' = \left\| [\mathbf{r}_k - \alpha_k \cdot \mathbf{d}_n + V_{rk}(t - \tau')] \alpha_k - \mathbf{d}_{cnk} + V_{ck}(t - \tau') \right\|. \quad (135)$$

Ignoring the cross-range terms in equation (135), viz.,  $\mathbf{d}_{cnk}$  and  $V_{ck}$ , one can make the following:

$$c\tau' = r_k - \alpha_k \cdot \mathbf{d}_n + V_{rk}(t - \tau'). \quad (136)$$

Solving for  $\tau'(t)$  yields

$$\tau'(t) = \frac{1}{c + V_{rk}} (r_k - \alpha_k \cdot \mathbf{d}_n + V_{rk}t). \quad (137a)$$

Using equations (132), (135) with  $V_{ck} \cong 0$ , and (137a), one can express  $\tau''(t)$  as

$$\tau''(t) = \frac{1}{c + V_{rk}} (r_k + \frac{V_{rk}}{c} \alpha_k \cdot \mathbf{d}_n + V_{rk}t). \quad (137b)$$

Then, from equations (130) and (137), it follows that (restoring the subscript where appropriate and the subscripts  $k, n$  on  $\tau$ )

$$t - \tau_{ink}(t) = \beta_{ik} \left[ t - \frac{1}{2} (1 + \beta_{ik}^{-1}) \tau_{ik} - \frac{1}{c} \alpha_{ik} \cdot \mathbf{d}_n \right], \quad (138)$$

where

$$\beta_{ik} = \frac{c - V_{irk}}{c + V_{irk}}, \quad (139)$$

and, as before,  $\tau_{ik} = 2r_k/c$ .

Equation (139) is, of course, the familiar expression for the Doppler effect of temporal dilation or contraction. In the second term on the right side of equation (138), assume that  $\beta_{ik} \approx 1$  then,

$$t - \tau_{ink}(t) = \beta_{ik}(t - \tau_{ik} - \frac{1}{c}\alpha_{ik} \cdot \mathbf{d}_n). \quad (140)$$

Now, for moving objects, the signal model becomes, from equations (129) and (140),

$$s_{in}(t) = \sum_{k=1}^{K_i} a_{ik} f[\beta_{ik}(t - \tau_{ink})], \quad T_1 \leq t \leq T_2, \quad (141)$$

where, as before (see equation (110)),

$$\tau_{ink} = \tau_{ik} - \frac{1}{c}\alpha_{ik} \cdot \mathbf{d}_n. \quad (142)$$

Define the  $N \times K_i$  matrix  $\mathbf{F}_i(t; \mathbf{V}_i)$  by

$$[\mathbf{F}_i(t; \mathbf{V}_i)]_{nk} = f[\beta_{ik}(t - \tau_{ink})], \quad T_1 \leq t \leq T_2, \quad (143)$$

where  $n = 1, 2, \dots, N$ , and  $k = 1, 2, \dots, K_i$ . Note that  $\mathbf{F}_i(t; \mathbf{0}) = \mathbf{F}_i(t)$ , as previously defined (see equation (14)). In vector form, equation (141) is

$$\mathbf{s}_i(t) = \mathbf{F}_i(t; \mathbf{V}_i) \mathbf{a}_i. \quad (144)$$

The results previously derived for the case of  $\mathbf{V}_i = \mathbf{0}$  ( $i = 1, 2, \dots, I$ ) can now be applied by substituting  $\mathbf{F}_i(t; \mathbf{V}_i)$  for  $\mathbf{F}_i(t)$ . The matrix  $\Phi_i$  now becomes  $\Phi_i(\mathbf{V}_i)$ , which is defined by

$$\Phi_i(\mathbf{V}_i) = \frac{1}{NE_f} \int_{T_1}^{T_2} \mathbf{F}_i^\dagger(t; \mathbf{V}_i) \mathbf{F}_i(t; \mathbf{V}_i) dt; \quad (145a)$$

and, of course,  $\Phi_i = \Phi_i(\mathbf{0})$ . Again using Parseval's theorem,  $\Phi_i(\mathbf{V}_i)$  can be written as

$$[\Phi_i(V_i)]_{jk} = (NE_d[\beta_{ij}][\beta_{ik}])^{-1} \int_{-\infty}^{\infty} f^*(\beta_{ij}^{-1}\omega) \sum_{n=1}^n \exp(i\omega(\beta_{ik}\tau_{ink} - \beta_{ij}\tau_{inj})) \frac{d\omega}{2\pi}. \quad (145b)$$

This equation shows explicitly how the Doppler effect, viz  $\beta_{ik}$ , figures into the spatial resolution (and ambiguity) of the processor. This case will be further developed in a sequel to this report. The expression for  $\Phi_i(V_i)$  can be interpreted in terms of the synthetic aperture effect. The effect on resolution of the processor can be characterized by examining the dependence of the variance ratio (equation (116)) on  $V_{ck}$  (and, of course, the other important parameters such as  $\Delta\theta$ ,  $\rho_N$ , etc). Moreover, the eigenanalysis shown in equations (111) and (112) can be modified to accommodate the synthetic aperture effect. Such analyses should reveal important interrelationships among object scattering characteristics (as modeled herein), waveform and array design, and motion induced effects within the unifying framework of the Bayes-optimal classifier.

#### NON-GAUSSIAN SIGNAL

To determine the Bayes-optimal classifier,  $x(t)$  was assumed to be a vector-valued Gaussian stochastic process. Recall that the model for  $x(t)$  is given by (under  $\mathcal{H}_i$ ,  $i = 1, 2, \dots, I$ )

$$x(t) = s_i(t) + n(t),$$

$$s_{in}(t) = \sum_{k=1}^{K_i} a_{ik} f(t - \tau_{ink}), \quad n = 1, 2, \dots, N,$$

and the noise process  $n(t)$  was assumed to be white and Gaussian. The scattering coefficients  $\{a_{ik}\}$  were not necessarily assumed to be Gaussian; however, the sum  $s_{in}(t)$  was assumed to be Gaussian, implicitly invoking the central limit theorem as is usually done. Suppose that the probability density function (pdf) of the random vector  $a_i$ , denoted by  $g_i(\cdot)$ , is now introduced explicitly and that it is not necessarily Gaussian. Since  $x$  is the discrete representation of the process  $x(t)$  as per equation (70), let the conditional pdf of  $x$  given  $a_i = A_i$  (and given  $\mathcal{H}_i$ ) be denoted by  $p(x|A_i; \mathcal{H}_i)$ , where  $A_i$  is some arbitrary realization of the random vector  $a_i$ . Then, the pdf of  $x$  given  $\mathcal{H}_i$  can be expressed as

$$p(x|\mathcal{H}_i) = \int p(x|A_i; \mathcal{H}_i) g_i(A_i) dA_i. \quad (146)$$

The likelihood ratio in equation (44) is now

$$\Lambda_i(\mathbf{x}) = \int \frac{p(\mathbf{x}|\mathbf{A}_i; \mathcal{H}_i)}{p(\mathbf{x}|\mathcal{H}_0)} g_i(\mathbf{A}_i) d\mathbf{A}_i; \quad (147)$$

or, as expected,

$$\Lambda_i(\mathbf{x}) = \int \Lambda_i(\mathbf{x}|\mathbf{A}_i) g_i(\mathbf{A}_i) d\mathbf{A}_i, \quad (148)$$

where  $\Lambda_i(\mathbf{x}|\mathbf{A}_i)$  is the conditional likelihood ratio defined by

$$\Lambda_i(\mathbf{x}|\mathbf{A}_i) = \frac{p(\mathbf{x}|\mathbf{A}_i; \mathcal{H}_i)}{p(\mathbf{x}|\mathcal{H}_0)}. \quad (149)$$

By equations (63) and (66), under  $\mathcal{H}_i$  and given  $\mathbf{a}_i = \mathbf{A}_i$ ,

$$\mathbf{x} = \mathbf{F}_i \mathbf{A}_i + \mathbf{n}, \quad (150)$$

where  $\mathbf{n}$  and  $\mathbf{F}_i$  are as previously defined (equations (53) and (68), respectively). Then,

$$\Lambda_i(\mathbf{x}|\mathbf{A}_i) = \frac{p_n(\mathbf{x} - \mathbf{F}_i \mathbf{A}_i)}{p_n(\mathbf{x})}, \quad (151)$$

where  $p_n(\cdot)$  is the pdf of  $\mathbf{n}$ . It is easy to show that if  $p_n(\cdot)$  is the Gaussian pdf, then the conditional likelihood ratio depends on  $\mathbf{x}$  only through  $\mathbf{y}_i = \mathbf{F}_i^\dagger \mathbf{K}_n^{-1} \mathbf{x}$ , the matched filter/beam-former operation (with prewhitening). This is so regardless of the particular form of  $g_i(\cdot)$ ; of course, the particular functional dependence of  $\Lambda_i(\cdot)$  on  $\mathbf{y}_i$  is determined by  $g_i(\cdot)$ , and is not, in general, quadratic as when  $\mathbf{a}_i$  is Gaussian. Future work will address alternatives to the Gaussian assumption with emphasis on the case of non-Gaussian stochastic signals and Gaussian noise. Much of the literature on the non-Gaussian (detection) problem focuses on the case of non-Gaussian noise and of a signal that is either known or stochastic. However, the non-Gaussian character of the signal is not explicitly characterized in the same way the noise is. The pdf of the signal is not explicitly introduced; usually, only its mean and covariance appear, as in the case of the locally optimal Bayes detectors (see references 25 and 26, for general treatment). Expansions including third and even higher order moments should be investigated, taking into account the important distinguishing features of the classification problem as formulated in this report. This

includes the possibility — a likely one in some important applications — that the *distribution* of the random signals will vary among alternative hypotheses  $\{\mathcal{H}_i\}$ .

## CONCLUSIONS

The active sonar classification problem can be approached as a likelihood ratio test of multiple, alternative hypotheses versus a noise-only null hypothesis. The data are, in general, vector-valued stochastic processes representing measurements from individual elements within a sonar array. Given an explicit form for the received signal model, which is statistically characterized for each alternative hypothesis (target class), explicit results can be derived for the likelihood ratio and various performance characteristics. Moreover, the optimal processor can be interpreted as an acoustic image processor. Generalizations of the results have been indicated and in some cases addressed in detail (e.g., the case of moving targets).

# APPENDIX A

## A CONTINUOUS MODEL FOR THE SIGNAL

In this appendix an optimal estimator of the random field  $a_i(\mathbf{r})$  is derived. The field  $a_i(\mathbf{r})$  appears in the signal model given by

$$s_i(t, \mathbf{d}) = \int_{\mathcal{R}_i} a_i(\mathbf{r}) f(t - \tau(\mathbf{r}, \mathbf{d})) d\mathbf{r}, \quad (\text{A-1})$$

where  $s_i(t, \mathbf{d})$  is the signal at position  $\mathbf{d}$  and time  $t$ ;  $f(t)$  is the transmitted waveform; the delay  $\tau(\mathbf{r}, \mathbf{d})$  for a non-refractive medium with sound speed  $c$  is given by

$$\tau(\mathbf{r}, \mathbf{d}) = \frac{1}{c} (\|\mathbf{r} - \mathbf{d}\| + r). \quad (\text{A-2})$$

The region  $\mathcal{R}_i$  contains the  $i^{\text{th}}$  object. Assume that  $s_i(t, \mathbf{d})$  is measured at  $N$  discrete sensor positions  $\mathbf{d}_n$ ,  $n = 1, 2, \dots, N$ , as is usual in practice. Then,

$$s_{in}(t) = \int_{\mathcal{R}_i} a_i(\mathbf{r}) f(t - \tau(\mathbf{r}, \mathbf{d}_n)) d\mathbf{r}, \quad (\text{A-3})$$

where  $s_{in}(t) = s_i(t, \mathbf{d}_n)$ ,  $n = 1, 2, \dots, N$ . When  $a(\mathbf{r}) = \sum_{i=1}^{K_i} a_i \delta(\mathbf{r} - \mathbf{r}_{ik})$ , equation (A-3) reduces to the discrete model of equation (11) with  $\tau_{ink} = \tau(\mathbf{r}_{ik}, \mathbf{d}_n)$ . The problem addressed here is the estimation of the field  $a_i(\mathbf{r})$  given the measurements  $\mathbf{x}(t)$ , where under the hypothesis  $\mathcal{H}_i$

$$\mathbf{x}(t) = \mathbf{s}_i(t) + \mathbf{n}(t), \quad (\text{A-4})$$

where  $\mathbf{s}(t) = [s_{i1}(t), s_{i2}(t), \dots, s_{iN}(t)]^T$ ,  $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_N(t)]^T$ , and  $\mathbf{n}(t) = [n_1(t), n_2(t), \dots, n_N(t)]^T$ , the measurement noise. (In the remainder of this appendix the subscript  $i$  will be deleted.)

Assume the following about the random field  $a(\mathbf{r})$ :

$$E[a(\mathbf{r})] = 0, \quad (\text{A-5})$$



and assume that the covariance  $K_a(r_1, r_2) = E[a(r_1)a^*(r_2)]$  is known for all  $r_1, r_2 \in \mathcal{R}$ . Regarding the noise, assume, for now, that  $E(n(t)) = 0$  for all  $t \in [T_1, T_2]$  and that  $K_n(t_1, t_2) = E(n(t_1)n^*(t_2))$  is known for all  $t_1, t_2 \in [T_1, T_2]$ ; moreover, assume that  $E[a(r)n^*(t)] = 0$ . Then, the following relations hold for  $E[x(t)]$  and  $K_x(t_1, t_2) = E[x(t_1)x^*(t_2)]$ :

$$E[x(t)] = 0, \quad (A-6)$$

and

$$K_x(t_1, t_2) = \int_{\mathcal{R}} \int_{\mathcal{R}} K_a(r_1, r_2) f(t_1, r_1) f^*(t_2, r_2) dr_1 dr_2 + K_n(t_1, t_2), \quad (A-7)$$

where  $f(t, r) = [f(t - \tau(r, d_1), f(t - \tau(r, d_2), \dots, f(t - \tau(r, d_N))]^T$ . Now, let  $\hat{a}(r)$  denote the best (minimum variance) linear, unbiased estimator of  $a(r)$  for all  $r \in \mathcal{R}$ ; thus,

$$E[\hat{a}(r)] = E[a(r)], \quad (A-8)$$

and the variance  $C_a(r)$  of the error  $\hat{a}(r) - a(r)$  is given by

$$C_a(r) = E[|\hat{a}(r) - a(r)|^2].$$

Since  $\hat{a}(r)$  is a linear estimator, it is of the form

$$\hat{a}(r) = \int h^T(r, t) x(t) dt, \quad (A-9)$$

where the vector-valued function  $h(r, t)$  is to be determined, and temporal integration is over  $[T_1, T_2]$ . From the orthogonal projection lemma

$$\text{cov}[(\hat{a}(r), (a(r) - \hat{a}(r))] = 0; \quad (A-10)$$

or, from equations (A-5) and (A-8),

$$E(|\hat{a}(r)|^2) = E[(a(r)\hat{a}^*(r))]. \quad (A-11)$$

Now from equation (A-9),

$$E[|\hat{a}(r)|^2] = \iint h^T(r, t_1) K_x(t_1, t_2) h^*(r, t_2) dt_1 dt_2, \quad (A-12)$$

and

$$E[a(r)\hat{a}^*(r)] = \int E[a(r)x^\dagger(t)h^*(r, t)]dt. \quad (A-13)$$

Combining equations (A-12) and (A-13) according to equation (A-11) yields

$$\int \{ E[a(r)x^\dagger(t_2)] - \int h^T(r, t_1) K_x(t_1, t_2) dt_1 \} h^*(r, t_2) dt_2 = 0 \quad (A-14)$$

identically for all  $r \in \mathcal{R}$ . Therefore, the expression within  $[\cdot]$  of equation (A-14) must be identically zero for  $h(r, t) \neq 0$ ; that is,

$$E[a(r)x^\dagger(t)] = \int h^T(r, t') K_x(t, t') dt', \quad (A-15)$$

the Wiener-Hopf equation.

Since

$$x(t) = \int_{\mathcal{R}} a(r)f(t, r)dr + n(t), \quad (A-16)$$

then

$$E[a(r)x^\dagger(t)] = \int_{\mathcal{R}} f^\dagger(t, r') K_a(r, r') dr'. \quad (A-17)$$

Substituting equations (A-7) and (A-17) into the Wiener-Hopf equation gives

$$\begin{aligned} \int_{\mathcal{R}} \mathbf{f}^{\dagger}(\mathbf{t}, \mathbf{r}') \mathbf{K}_a(\mathbf{r}, \mathbf{r}') d\mathbf{r}' &= \int_{\mathcal{R}} \int_{\mathcal{R}} \int_{\mathcal{R}} \mathbf{h}^{\dagger}(\mathbf{r}, \mathbf{t}') \mathbf{f}(\mathbf{t}', \mathbf{r}') d\mathbf{t}' \mathbf{K}_a(\mathbf{r}', \mathbf{r}'') \mathbf{f}^{\dagger}(\mathbf{t}, \mathbf{r}'') d\mathbf{r}' d\mathbf{r}'' \\ &+ \int \mathbf{h}^{\dagger}(\mathbf{r}, \mathbf{t}') \mathbf{K}_n(\mathbf{t}', \mathbf{t}) d\mathbf{t}'. \end{aligned} \quad (\text{A-18})$$

Now, define a function  $\gamma(\mathbf{r}_1, \mathbf{r}_2)$  over  $\mathcal{R} \times \mathcal{R}$  such that

$$\mathbf{h}^{\dagger}(\mathbf{r}, \mathbf{t}) = \int_{\mathcal{R}} \gamma(\mathbf{r}, \mathbf{r}') \mathbf{f}^{\dagger}(\mathbf{t}, \mathbf{r}') d\mathbf{r}'. \quad (\text{A-19})$$

Then, equation (A-18) becomes

$$\begin{aligned} \int_{\mathcal{R}} \mathbf{f}^{\dagger}(\mathbf{t}, \mathbf{r}') \mathbf{K}_a(\mathbf{r}, \mathbf{r}') d\mathbf{r}' &= \int_{\mathcal{R}} \int_{\mathcal{R}} \int_{\mathcal{R}} \gamma(\mathbf{r}, \mathbf{r}''') \mathbf{f}^{\dagger}(\mathbf{t}', \mathbf{r}''') \mathbf{f}(\mathbf{t}', \mathbf{r}') d\mathbf{t}' \mathbf{K}_a(\mathbf{r}', \mathbf{r}'') \mathbf{f}^{\dagger}(\mathbf{t}, \mathbf{r}'') d\mathbf{r}' d\mathbf{r}'' d\mathbf{r}''' \\ &+ \int \int_{\mathcal{R}} \gamma(\mathbf{r}, \mathbf{r}') \mathbf{f}^{\dagger}(\mathbf{t}', \mathbf{r}') \mathbf{K}_n(\mathbf{t}', \mathbf{t}) d\mathbf{t}' d\mathbf{r}'. \end{aligned} \quad (\text{A-20})$$

Defining the signal (spatial) correlation function  $\phi(\mathbf{r}_1, \mathbf{r}_2)$  by

$$\phi(\mathbf{r}_1, \mathbf{r}_2) = \int \mathbf{f}^{\dagger}(\mathbf{t}, \mathbf{r}_1) \mathbf{f}(\mathbf{t}, \mathbf{r}_2) d\mathbf{t} \quad (\text{A-21})$$

makes equation (A-20) become

$$\begin{aligned} \int_{\mathcal{R}} \mathbf{f}^{\dagger}(\mathbf{t}, \mathbf{r}') \mathbf{K}_a(\mathbf{r}, \mathbf{r}') d\mathbf{r}' &= \int_{\mathcal{R}} \int_{\mathcal{R}} \int_{\mathcal{R}} \gamma(\mathbf{r}, \mathbf{r}''') \phi(\mathbf{r}''', \mathbf{r}') \mathbf{K}_a(\mathbf{r}', \mathbf{r}'') \mathbf{f}^{\dagger}(\mathbf{t}, \mathbf{r}'') d\mathbf{r}' d\mathbf{r}'' d\mathbf{r}''' \\ &+ \frac{N_0}{2} \int_{\mathcal{R}} \gamma(\mathbf{r}, \mathbf{r}') \mathbf{f}^{\dagger}(\mathbf{t}, \mathbf{r}') d\mathbf{r}', \end{aligned} \quad (\text{A-22})$$

where it is now assumed that  $K_n(t_1, t_2) = \frac{N_0}{2} \delta(t_1 - t_2) \mathbf{I}$ , a white noise process. After rearranging terms, one can write equation (A-22) as

$$\int_{\mathcal{R}} \left[ K_a(\mathbf{r}, \mathbf{r}') - \int_{\mathcal{R}} \int_{\mathcal{R}} \gamma(\mathbf{r}, \mathbf{r}'') \phi(\mathbf{r}'', \mathbf{r}') K_a(\mathbf{r}', \mathbf{r}'') d\mathbf{r}'' d\mathbf{r}''' - \frac{N_0}{2} \gamma(\mathbf{r}, \mathbf{r}') \right] \mathbf{f}^T(t, \mathbf{r}'') d\mathbf{r}'' = 0. \quad (\text{A-23})$$

Since  $\mathbf{f}(t, \mathbf{r}) \neq 0$  for all  $t \in [T_1, T_2]$  and  $\mathbf{r} \in \mathcal{R}$ , then the expression within  $[\cdot]$  in equation (A-23) is identically zero; that is,

$$\int_{\mathcal{R}} \gamma(\mathbf{r}_1, \mathbf{r}') \int_{\mathcal{R}} \phi(\mathbf{r}', \mathbf{r}'') K_a(\mathbf{r}'', \mathbf{r}_2) d\mathbf{r}'' d\mathbf{r}' + \frac{N_0}{2} \gamma(\mathbf{r}_1, \mathbf{r}_2) = K_a(\mathbf{r}_1, \mathbf{r}_2). \quad (\text{A-24})$$

Define the function  $\beta(\mathbf{r}_1, \mathbf{r}_2)$  over  $\mathcal{R} \times \mathcal{R}$  by

$$\beta(\mathbf{r}_1, \mathbf{r}_2) = \int_{\mathcal{R}} \phi(\mathbf{r}_1, \mathbf{r}') K_a(\mathbf{r}', \mathbf{r}_2) d\mathbf{r}'; \quad (\text{A-25})$$

then, equation (A-24) can be written as

$$\int_{\mathcal{R}} \gamma(\mathbf{r}_1, \mathbf{r}') \beta(\mathbf{r}', \mathbf{r}_2) d\mathbf{r}' + \frac{N_0}{2} \gamma(\mathbf{r}_1, \mathbf{r}_2) = K_a(\mathbf{r}_1, \mathbf{r}_2). \quad (\text{A-26})$$

In summary, the minimum variance, linear, unbiased estimator of the random field  $\mathbf{a}(\mathbf{r})$  is given by, combining equations (A-9) and (A-19),

$$\hat{\mathbf{a}}(\mathbf{r}) = \int_{\mathcal{R}} \gamma(\mathbf{r}, \mathbf{r}') \int \mathbf{f}^T(t, \mathbf{r}') \mathbf{x}(t) dt d\mathbf{r}', \quad (\text{A-27})$$

where the kernel  $\gamma(\mathbf{r}_1, \mathbf{r}_2)$  is the solution to the integral equation (A-26).

To complete this treatment of the continuous model an expression for the error variance  $C_a(\mathbf{r}) = E[\{\hat{\mathbf{a}}(\mathbf{r}) - \mathbf{a}(\mathbf{r})\}^2]$  is derived. Now,  $C_a(\mathbf{r})$  can be written as

$$C_a(\mathbf{r}) = E\{\mathbf{a}^*(\mathbf{r})[\mathbf{a}(\mathbf{r}) - \hat{\mathbf{a}}(\mathbf{r})]\} - E\{\hat{\mathbf{a}}^*(\mathbf{r})[\mathbf{a}(\mathbf{r}) - \hat{\mathbf{a}}(\mathbf{r})]\}. \quad (\text{A-28})$$

The second term on the right of equation (A-28) vanishes because of the orthogonal projection lemma, equation (A-10). Next, by substituting equation (A-16) into equation (A-27),  $\hat{a}(\mathbf{r})$  becomes

$$\begin{aligned}\hat{a}(\mathbf{r}) = & \int_{\mathcal{R}} \int_{\mathcal{R}} \gamma(\mathbf{r}, \mathbf{r}') \phi(\mathbf{r}', \mathbf{r}'') a(\mathbf{r}'') d\mathbf{r}' d\mathbf{r}'' \\ & + \int_{\mathcal{R}} \gamma(\mathbf{r}, \mathbf{r}') f^\dagger(t, \mathbf{r}') \mathbf{n}(t) d\mathbf{r}' dt.\end{aligned}\quad (\text{A-29})$$

From equation (A-29),  $E\{a^*(\mathbf{r})[a(\mathbf{r}) - \hat{a}(\mathbf{r})]\}$  becomes

$$E\{a^*(\mathbf{r})[a(\mathbf{r}) - \hat{a}(\mathbf{r})]\} = K_a(\mathbf{r}, \mathbf{r}) - \int_{\mathcal{R}} \int_{\mathcal{R}} \gamma(\mathbf{r}, \mathbf{r}') \phi(\mathbf{r}', \mathbf{r}'') K_a(\mathbf{r}'', \mathbf{r}) d\mathbf{r}' d\mathbf{r}''.\quad (\text{A-30})$$

The second term on the right of equation (A-29) is uncorrelated with  $a^*(\mathbf{r})$  since  $E[(a^*(\mathbf{r})\mathbf{n}(t))] = 0$ .

The expression for  $C_a(\mathbf{r})$  becomes, using the definition of  $\beta(\mathbf{r}_1, \mathbf{r}_2)$  in equation (A-25),

$$C_a(\mathbf{r}) = K_a(\mathbf{r}, \mathbf{r}) - \int_{\mathcal{R}} \gamma(\mathbf{r}, \mathbf{r}') \beta(\mathbf{r}', \mathbf{r}) d\mathbf{r}',\quad (\text{A-31})$$

which finally becomes, using equation (A-26),

$$C_a(\mathbf{r}) = \frac{N_0}{2} \gamma(\mathbf{r}, \mathbf{r}),\quad (\text{A-32})$$

where, as before,  $\gamma(\mathbf{r}_1, \mathbf{r}_2)$  is the solution to the integral equation (A-26).

## APPENDIX B

### PERFORMANCE FOR THE BINARY CLASSIFICATION PROBLEM

#### BACKGROUND

This appendix addresses the problem of determining the performance of the optimum processor for the binary classification problem described in the main text. The intent is to first present the details of classification performance prediction for the general binary case. It is shown that analytic solutions to classification performance prediction are difficult to obtain for general target geometries and array/waveform configurations; therefore, simplifying assumptions are made so that analytic solutions can be derived. These are used to illustrate fairly simple but instructive cases.

The conditions of the problem can be briefly restated as follows: the test region is assumed to be identical under each hypothesis; each scattering coefficient covariance matrix is assumed to have full rank, and the noise is assumed to be both temporally and spatially white. For the binary problem there are two hypotheses:  $\mathcal{H}_1$  and  $\mathcal{H}_2$ . Under these conditions, the sufficient statistic for deciding whether the received signal was from object 1 or 2 was shown in equation (100) to be

$$L = \mathbf{y}^\dagger (\mathbf{H}_1 - \mathbf{H}_2) \mathbf{y}, \quad (\text{B-1})$$

where

$$\mathbf{y} = \frac{2}{N_o} \int_{T_1}^{T_2} \mathbf{F}^\dagger(t) \mathbf{x}(t) dt, \quad (\text{B-2})$$

and

$$\mathbf{H}_i = \mathbf{K}_{a_i} [\mathbf{I} + \rho_N \Phi \mathbf{K}_{a_i}]^{-1} = [\mathbf{K}_{a_i}^{-1} + \rho_N \Phi]^{-1}, \quad i = 1, 2. \quad (\text{B-3})$$

## PERFORMANCE MEASURE

Classification performance is a function of the distribution of the scalar random variable  $L$  under  $\mathcal{H}_1$  or  $\mathcal{H}_2$ . If  $P_{ij}$  is defined to be the probability of selecting object  $i$  when object  $j$  occurred, then

$$P_{11}(\eta) = \int_{\eta}^{\infty} p(L|\mathcal{H}_1) dL, \quad (\text{B-4a})$$

$$P_{22}(\eta) = \int_{-\infty}^{\eta} p(L|\mathcal{H}_2) dL, \quad (\text{B-4b})$$

and

$$P_{12}(\eta) = \int_{\eta}^{\infty} p(L|\mathcal{H}_2) dL, \quad (\text{B-5a})$$

$$P_{21}(\eta) = \int_{-\infty}^{\eta} p(L|\mathcal{H}_1) dL, \quad (\text{B-5b})$$

where  $p(L|\mathcal{H}_i)$  is the probability density function of  $L$  conditioned on the  $i^{\text{th}}$  hypothesis ( $i = 1, 2$ ), and  $\eta$  is the decision threshold.  $P_{11}$  and  $P_{22}$  are measures of correct classification, while  $P_{12}$  and  $P_{21}$  yield measures of incorrect classification.

### General Matrices

Let us determine the density functions  $p(L|\mathcal{H}_i)$ ,  $i = 1, 2$ . Note that  $y$  is a linear functional of the vector-valued, complex Gaussian process  $x(t)$ ; therefore, it is also a complex Gaussian vector with

$$E[y|\mathcal{H}_i] = 0, \quad i = 1, 2, \quad (\text{B-6})$$

and

$$\mathbf{K}_{y_i} \equiv E [\mathbf{y} \mathbf{y}^\dagger | \mathcal{H}_i] = \rho_N^2 \Phi \mathbf{K}_{a_i} \Phi^\dagger + \rho_N \Phi, \quad i = 1, 2, \quad (\text{B-7})$$

where  $E [\cdot]$  denotes expectation.

$L$  is therefore a complex quadratic form in the Gaussian vector  $\mathbf{y}$  (note that the matrix  $\mathbf{H}_1 - \mathbf{H}_2$  is Hermetian but not necessarily positive definite). The characteristic function of quadratic forms in complex Gaussian vectors is well known and for the current problem it can be shown to be (reference 23)

$$M_i(j\omega) \equiv E [\exp(j\omega L) | \mathcal{H}_i] = \{ \det [\mathbf{C}_i(j\omega)] \}^{-1}, \quad i = 1, 2, \quad (\text{B-8a})$$

where

$$\mathbf{C}_i(j\omega) = \mathbf{I} - j\omega \mathbf{K}_{y_i} (\mathbf{H}_1 - \mathbf{H}_2), \quad i = 1, 2, \quad (\text{B-8b})$$

and  $\det [\cdot]$  is the determinant.

Moreover, it can be shown that (reference 25)

$$M_i(j\omega) = \prod_{k=1}^K (1 - j\omega \zeta_{ki})^{-1}, \quad (\text{B-8c})$$

where  $\zeta_{ki}$  is the  $k^{\text{th}}$  eigenvalue of the matrix  $(\mathbf{H}_1 - \mathbf{H}_2) \mathbf{K}_{y_i}$ ,  $i = 1, 2$ . The probability density functions of interest can then be written as

$$p(L | \mathcal{H}_i) = \frac{1}{2\pi} \int_{-\infty}^{\infty} M_i(j\omega) \exp(-j\omega L) d\omega, \quad i = 1, 2. \quad (\text{B-9})$$

### Diagonal Matrices

In general, considering the complexity of the matrices  $\mathbf{H}_i$  ( $i = 1, 2$ ), any performance evaluation based on equation (B-8) would have to be made on a per case basis by numerical means and would not necessarily yield any insight into the classification performance of the processor. If



the following simplifying assumptions are made, some insight into processor performance may be gained. First, assume that the waveform and the array are capable of perfect resolution in range and angle (for the particular test region assumed). That is, assume

$$\Phi = I. \quad (B-10)$$

Also, assume that under each hypothesis, the  $K$  scatterers in the test volume are statistically independent so that  $K_{ai}$  now represents a discrete version of the conventional scattering function (reference 2). Thus,  $K_{ai}$  is a diagonal matrix ( $i = 1, 2$ ). Given these assumptions,  $H_i$ ,  $K_{yi}$ , and therefore  $C_i(j\omega)$  can be shown to be diagonal matrices and equation (B-8) can then be written as

$$M_i(j\omega) = \prod_{k=1}^K \frac{1}{[C_i(j\omega)]_{kk}}, \quad i = 1, 2, \quad (B-11)$$

where  $[C_i(j\omega)]_{kk}$  is the  $k^{\text{th}}$  diagonal element of  $C_i(j\omega)$ . Thus, under each hypothesis,  $L$  consists of the sum of  $K$  independent random variables

$$L = \sum_{k=1}^K l_k,$$

where

$$l_k = \left( \frac{\sigma_{k1}^2}{1 + \sigma_{k1}^2 \rho_N} - \frac{\sigma_{k2}^2}{1 + \sigma_{k2}^2 \rho_N} \right) y_k^2.$$

Here  $\rho_N$  is the input signal-to-noise ratio  $= 2NE_f/N_o$ , and  $\sigma_{ki}^2$  is the  $k^{\text{th}}$  diagonal element of  $K_{ai}$  (scattering strength of the  $k^{\text{th}}$  cell under the  $i^{\text{th}}$  hypothesis).

The density functions for  $l_k$  ( $k = 1, 2, \dots, K$ ) will now be determined. If a generalized receive signal-to-noise ratio  $d_{ki} = \rho_N \sigma_{ki}^2 = 2 \sigma_{ki}^2 NE_f/N_o$  is defined for each cell then

$$\frac{1}{[C_1(j\omega)]_{kk}} = \frac{1}{(1 - j\omega\alpha_{k1})}, \quad (B-12a)$$

where

$$\alpha_{k1} = \frac{d_{k1} - d_{k2}}{d_{k2} + 1}, \quad (\text{B-12b})$$

and

$$\frac{1}{[C_2(j\omega)]_{kk}} = \frac{1}{(1 - j\omega\alpha_{k2})} \quad (\text{B-13a})$$

where

$$\alpha_{k2} = \frac{d_{k1} - d_{k2}}{d_{k1} + 1} \quad (\text{B-13b})$$

Therefore, each of the K components  $l_k$  of L is distributed as follows:

$$\begin{aligned} p(l_k | \mathcal{H}_i) &\equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} [C_i(j\omega)]_{kk} \exp(-j\omega l_k) d\omega = \frac{1}{|\alpha_{ki}|} \exp\left(-\frac{l_k}{\alpha_{ki}}\right), & \frac{l_k}{\alpha_{ki}} \geq 0, \\ &= 0, & \frac{l_k}{\alpha_{ki}} < 0. \end{aligned} \quad (\text{B-14})$$

Under each of the two hypotheses, L is equal to the sum and/or difference (note  $\alpha_{ki}$  can be positive or negative) of independent, not necessarily identically distributed, exponential random variables. Its density functions  $p(L | \mathcal{H}_i)$  ( $i = 1, 2$ ) can be found from equation (B-9) via residue theory. These functions can be used to solve for  $P_{ij}$  ( $i, j = 1, 2$ ) of equations (B-4) and (B-5).

Some examples will now be investigated by specifying values for the scattering strengths  $\sigma_{ki}^2$  or, equivalently, the generalized receive signal-to-noise ratios  $d_{ki}$  ( $k = 1, 2, \dots, K$  and  $i = 1, 2$ ).

## PERFORMANCE EXAMPLES

Before addressing specific target descriptions, some additional points should be made. The test volume is assumed to be identical for each hypothesis, and it comprises  $K$  cells with an independent specular scatterer in each cell. If under a specific hypothesis no target is present, and the scattered return is due to reverberation alone, all cells in the test volume will possess uniform scattering strength. If, in addition to reverberation, a target is present under a specific hypothesis, the uniform scattering strength of a number of the cells ( $\leq K$ ) in the test volume may be replaced with new scattering strengths; the number of cells, their location, and the magnitude of their scattering strengths will vary according to the target. Thus, it is assumed that each individual scatterer falls into one of two categories: target-like or reverberation, and the difference between hypotheses  $\mathcal{H}_1$  and  $\mathcal{H}_2$  is due to the number, location, and scattering strength of the target-like scatterers. Thus, although their number may be different, it is assumed in these examples that reverberation scatterers are of equal strength under both hypotheses.

Because perfect resolution has been assumed, it is possible to index the  $K$  cells arbitrarily, but identically, under both  $\mathcal{H}_1$  and  $\mathcal{H}_2$ . Thus, the following convention will be adopted: from the  $K$  test cells, find the total  $J$  such that  $J \leq K$  and,

$$d_{j1} \neq d_{j2}, \quad j = 1, 2, \dots, J. \quad (\text{B-15})$$

The remaining  $K - J$  cells of equal scattering strength do not contribute to either equation (B-12) or (B-13) and, therefore, have no effect on the characteristic function in equation (B-11). The specific ordering of the  $J$  cells with unequal scattering strengths can be arbitrary as mentioned before. In light of this, the intersection of reverberation cells under  $\mathcal{H}_1$  and  $\mathcal{H}_2$  does not enter into the problem; this is also the case for the intersection of any target-like cells of equal strength.

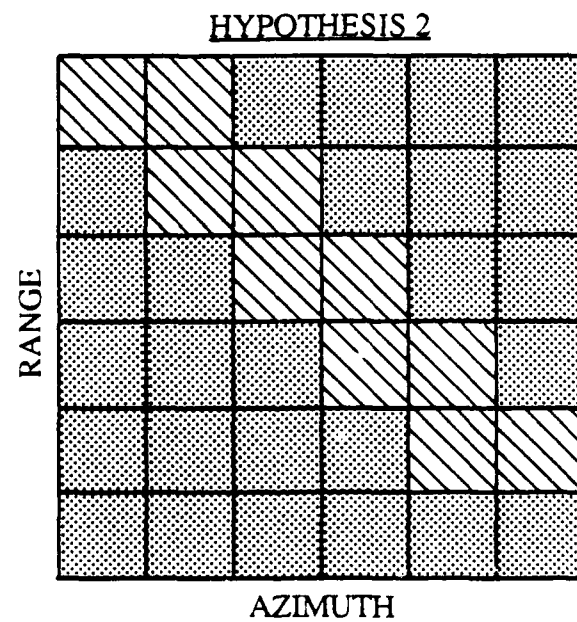
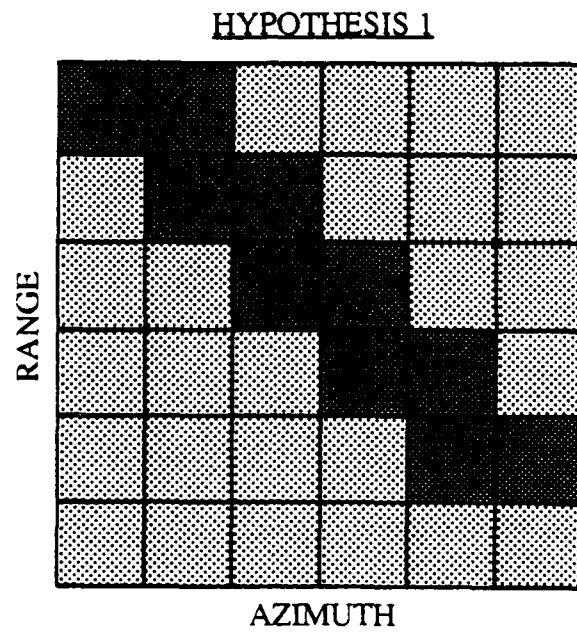
Two examples will now be examined. See table B-1 for a summary of the target descriptions.

Table B-1. Target Descriptions

$\mathcal{H}_1$	$\mathcal{H}_2$	Comments
Case 1. J target cells $d_{k1} = \beta, k = 1, 2, \dots, J$ K - J reverberation cells $d_{k1} = \gamma, k = J + 1, \dots, K$	J target cells $d_{k2} = \Psi, k = 1, 2, \dots, J$ K - J reverberation cells $d_{k2} = \gamma, k = J + 1, \dots, K$	Assume $\beta > \Psi$ . Targets collocated under $\mathcal{H}_1, \mathcal{H}_2$ . K - J reverberation cells do not affect performance
Case 2. $M < J$ target cells $d_{k1} = \beta, k = 1, 2, \dots, M$ K - M reverberation cells $d_{k1} = \gamma, k = J + 1, \dots, K$	J target cells $d_{k2} = \Psi, k = 1, 2, \dots, J$ K - J reverberation cells $d_{k1} = \gamma, k = J + 1, \dots, K$	Assume $M\beta = J\Psi$ and $\gamma < \Psi < \beta$ . K - J reverberation cells do not affect performance, but $J - M + 1$ reverberation cells do.

### Case 1

The target is located in the exact same J cells under each hypothesis, and each target is assumed to have uniform scattering strength (see figure B-1a for an example of the test volume geometry). Under these conditions, performance is a function of the number of target cells J and their relative strengths  $\beta$  and  $\psi$ ; the remaining K - J cells that correspond to reverberation do not affect performance (see equations (B-12b) and (B-13b)).






-  REVERBERATION CELL ( $\gamma$ )
-  TARGET CELL  $H_1$  ( $\beta$ )
-  TARGET CELL  $H_2$  ( $\psi$ )

Figure B-1a. Example of a Two-dimensional Test Volume for Case 1;  $K = 36$  Test Cells;  
 $J = 10$  Target Cells (Note: Target cell locations identical under each hypothesis.)

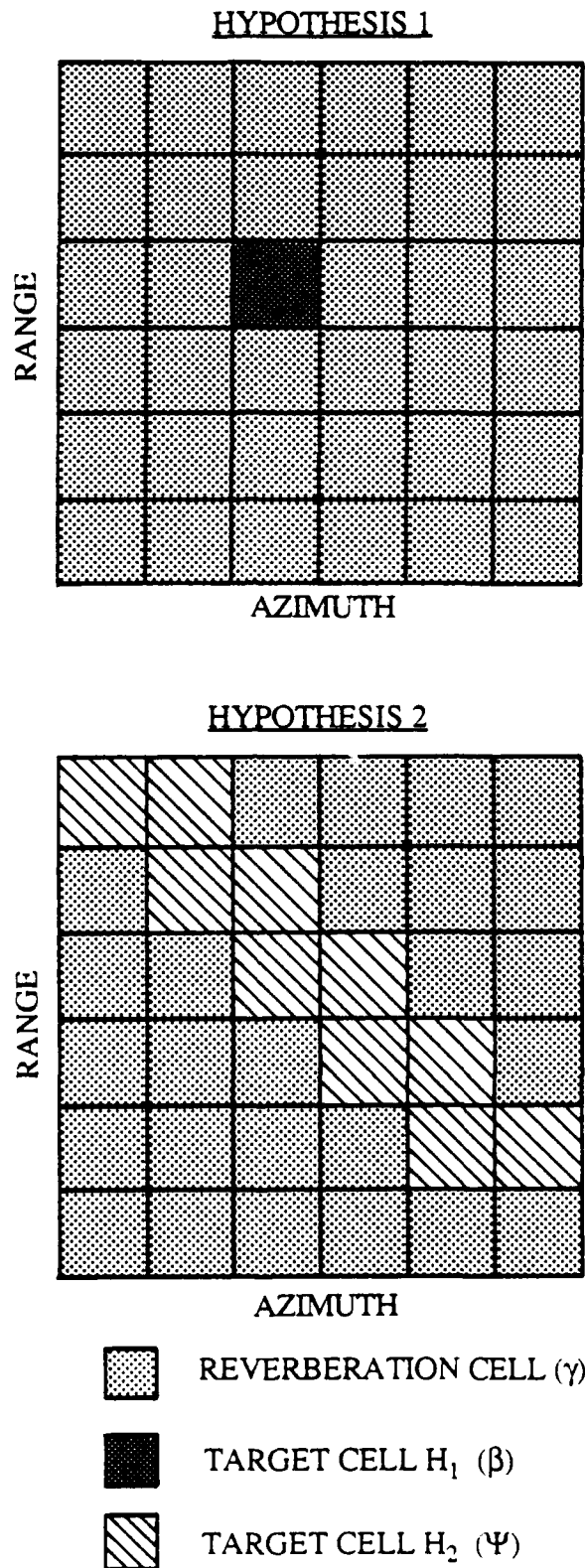


Figure B-1b. Example of a Two-dimensional Test Volume for Case 1;  $K = 36$  Test Cells;  $M = 1$  Target Cell Under  $H_1$ ;  $J = 10$  Target Cells Under  $H_2$  (Note: Target cell locations under  $H_1$  are a subset of those under  $H_2$ .)

$$\begin{aligned}\alpha_{k1} &= \frac{\beta - \psi}{\psi + 1} \equiv \kappa, & k = 1, 2, \dots, J, \\ &= 0, & k = J + 1, J + 2, \dots, K,\end{aligned}\tag{B-16a}$$

$$\begin{aligned}\alpha_{k2} &= \frac{\beta - \psi}{\beta + 1} \equiv \nu, & k = 1, 2, \dots, J, \\ &= 0, & k = J + 1, J + 2, \dots, K.\end{aligned}\tag{B-16b}$$

Since  $\beta > \psi$ ,  $\kappa$  and  $\nu$  are both positive. Thus, under each hypothesis,  $L$  follows a gamma distribution:

$$\begin{aligned}p(L|\mathcal{H}_1) &= \frac{(L)^{J-1}}{\kappa^J(J-1)!} \exp\left(-\frac{L}{\kappa}\right), & \text{for } L \geq 0, \\ &= 0, & \text{for } L < 0,\end{aligned}\tag{B-17a}$$

and

$$\begin{aligned}p(L|\mathcal{H}_2) &= \frac{(L)^{J-1}}{\nu^J(J-1)!} \exp\left(-\frac{L}{\nu}\right), & \text{for } L \geq 0, \\ &= 0, & \text{for } L < 0.\end{aligned}\tag{B-17b}$$

This yields the following expressions for classifier performance.

$$P_{11}(\eta) = \exp\left(-\frac{\eta}{\kappa}\right) \sum_{k=0}^{J-1} \frac{\eta^k}{k! \kappa^k},\tag{B-18a}$$

$$P_{12}(\eta) = \exp\left(-\frac{\eta}{\nu}\right) \sum_{k=0}^{J-1} \frac{\eta^k}{k! \nu^k} = \exp\left(-\frac{\eta(1+\kappa)}{\kappa}\right) \sum_{k=0}^{J-1} \frac{\eta^k (1+\kappa)^k}{k! \kappa^k},\tag{B-18b}$$

since  $\nu = \kappa/(1+\kappa)$ .

Figure B-2 shows the performance curve ( $P_{11}$  versus  $P_{12}$ ) for some representative values of  $\beta$ ,  $\psi$ , and  $J$ . In this example, both the scattering strength ratio ( $\beta/\psi$ ) and total target receive signal-to-noise ratios ( $J\beta$  and  $J\psi$ ) are held constant as the number of cells encompassed by the target is increased; therefore, the individual cell receive signal-to-noise ratios ( $\beta$  and  $\psi$ ) decrease. Each target in this example can be thought of as belonging to a class of targets characterized by constant target receive signal-to-noise ratio ( $J\beta$  or  $J\psi$ ).

The results indicate that, for a given  $P_{12}$  (probability of incorrect classification), there is a particular set of targets (characterized by  $J'$ ,  $\beta'$  and  $J'$ ,  $\psi'$ ), taken from the two classes, that are most easily distinguished from each other. This is also represented in figure B-3, where  $P_{11}$  (probability of correct classification) is plotted as a function of the number of cells  $J$  for a fixed  $P_{12}$ . A family of curves, parameterized by fixed scattering strength ratio  $\beta/\psi$  is shown. This plot illustrates that there is a partitioning of the independent target scattering strengths into a particular number of cells, for which performance is best. This result, spreading of the total fixed receive energy over a particular number of independent channels (target cells) for maximum performance, is an example of the well-known "diversity" phenomenon (reference 2). Of course, in the present example, the system designer has no control over the target physics so that the result indicates only which set of targets within the class are most easily distinguished from each other.

As a second example of Case 1, let the number of target cells under each hypothesis  $J$  and the total target receive signal-to-noise ratio under  $\mathcal{H}_2$ ,  $J\psi$  remain fixed. Let the individual cell receive signal-to-noise ratio under  $\mathcal{H}_1$ ,  $\beta$  increase. As in the first example, the remaining  $K - J$  reverberation cells do not affect performance.

As expected, performance improves with increased  $\beta$ . The results, shown in figure B-4, have a number of interesting interpretations. First, as  $\beta$  is increased, the performance gain can be viewed as the result of increasing the total target receive signal-to-noise ratio under  $\mathcal{H}_1$  relative to that under  $\mathcal{H}_2$ . Second, assume that  $\psi = \gamma$ , that is, assume that hypothesis 2 represents reverberation only; all  $K$  cells under  $\mathcal{H}_2$  have a receive signal-to-noise ratio of  $\psi$ . Although  $K - J$  reverberation cells do not affect performance,  $J$  of the cells now do. As  $\beta$  is increased, the performance gain can be viewed as increasing the total target receive signal-to-noise ratio under  $\mathcal{H}_1$  relative to the total receive signal-to-noise ratio of  $J$  of the reverberation cells under  $\mathcal{H}_2$ . Third, assume that  $\beta = \gamma$ , that is, assume that hypothesis 1 represents reverberation only. Recall that  $\beta > \psi$ , so that in this interpretation greater performance gains are made as the total receive signal-to-noise ratio of the reverberation increases beyond that of the target. This interpretation could have relevance to target strength reduction efforts.



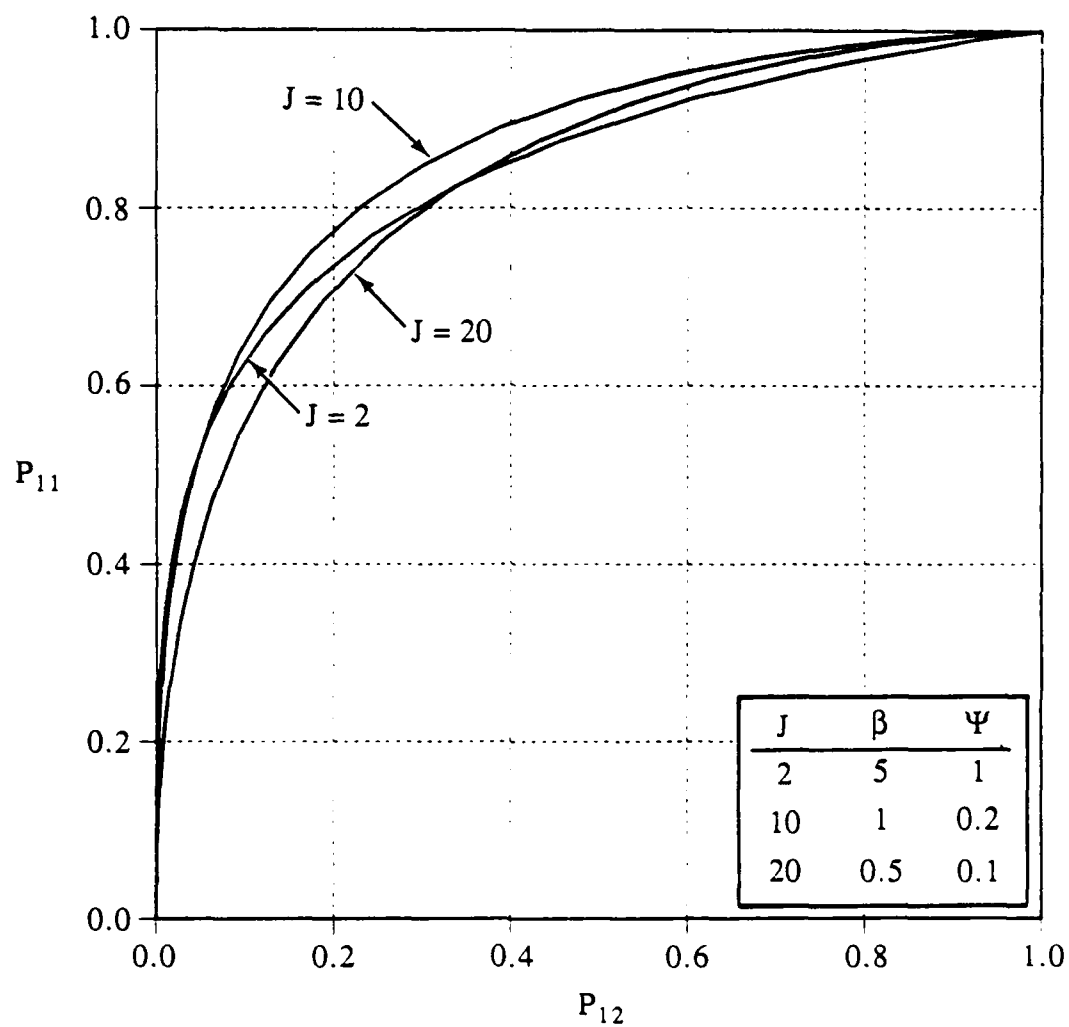


Figure B-2. Receiver Operating Characteristic (ROC) Curve for Case 1;  
 $J\beta = 10$ ;  $J\Psi = 2$ ;  $\beta/\Psi = 5$ ; Vary  $J = 2, 10, 20$

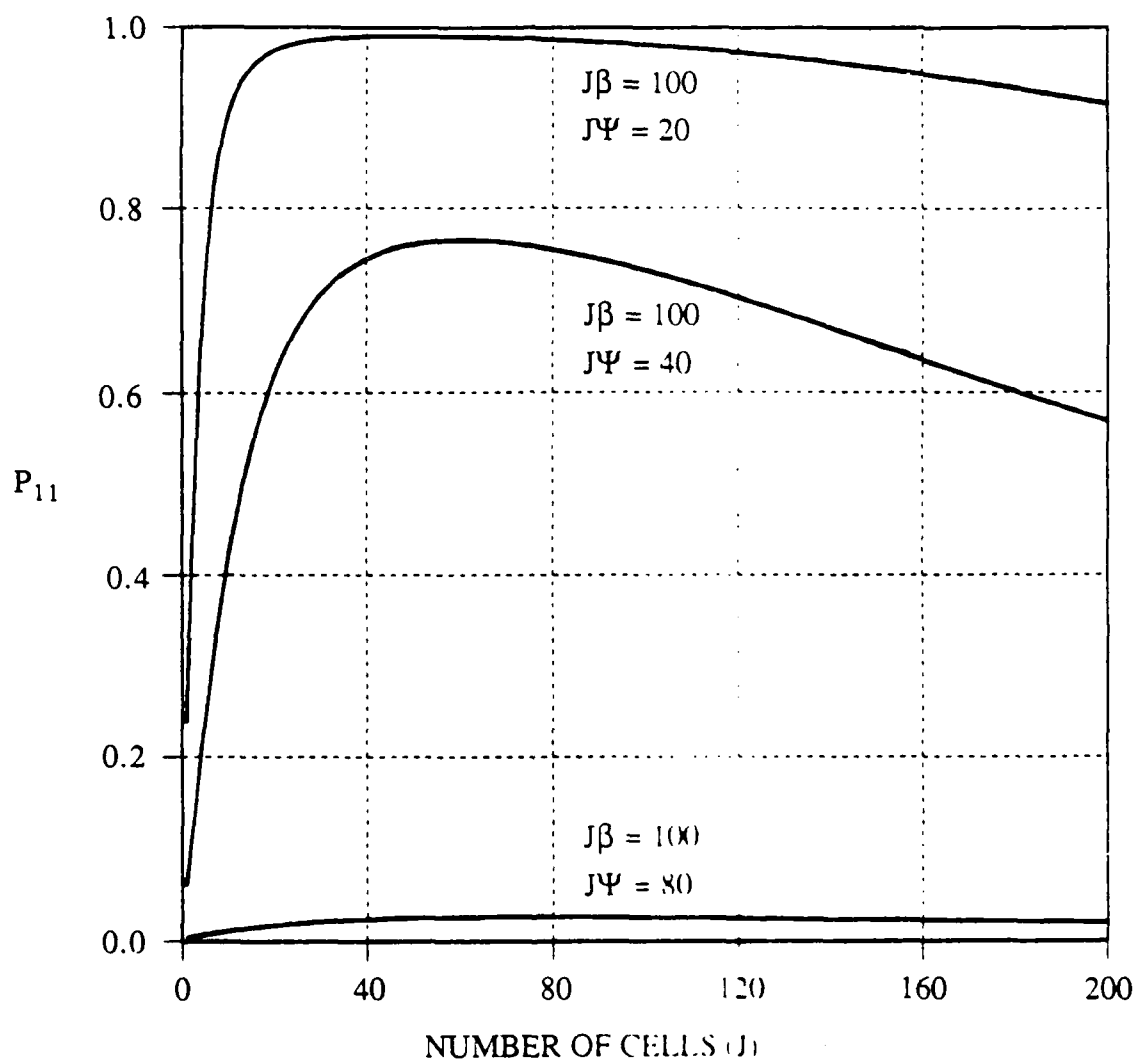


Figure B-3.  $P_{11}$  as a Function of  $J$  for Case 1 ( $P_{12} = 0.001$ )

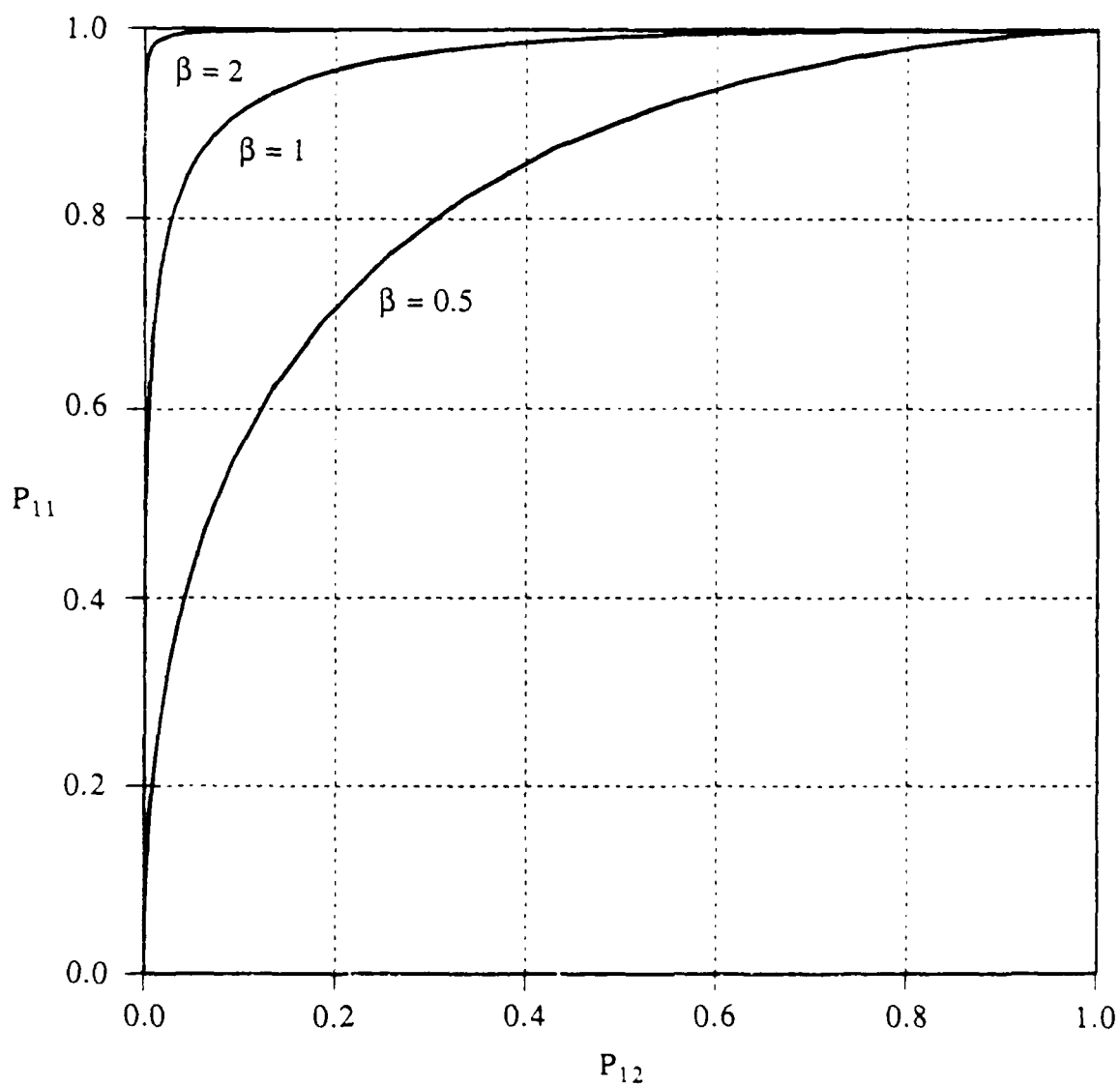


Figure B-4. Receiver Operating Characteristic (ROC) Curve for Case 1;  
 $J = 20$ ;  $\Psi = 0.1$ ; vary  $\beta = 0.5, 1, 2$

## Case 2

As in example 1, each target comprises scatterers of constant strength. Under hypothesis 1, the target consists of  $M$  cells. Under hypothesis 2, the target comprises  $J > M$  cells, the same  $M$  cells as under hypothesis 1, and  $J - M$  additional cells (see figure B-1b for an example of the test volume geometry). In addition, it is assumed that the total target scattering strength under each hypothesis is equal ( $M\beta = J\psi$ ). Because of the unequal number of target scatterers, a number only  $J - M$  of reverberation scatterers must be accounted for. The results are as follows:

$$\alpha_{k1} = \frac{\beta - \psi}{\psi + 1} \equiv \kappa, \quad k = 1, 2, \dots, M, \quad (\text{B-19a})$$

$$\alpha_{k1} = \frac{\gamma - \psi}{\psi + 1} \equiv \lambda, \quad k = M + 1, M + 2, \dots, J, \quad (\text{B-19b})$$

and

$$\alpha_{k2} = \frac{\beta - \psi}{\beta + 1} \equiv \nu, \quad k = 1, 2, \dots, M, \quad (\text{B-19c})$$

$$\alpha_{k2} = \frac{\gamma - \psi}{\gamma + 1} \equiv \mu, \quad k = M + 1, M + 2, \dots, J. \quad (\text{B-19d})$$

Since  $\gamma < \psi < \beta$ , both  $\lambda$  and  $\mu$  are negative resulting in  $L$  (under each hypothesis) being equivalent to the difference of two independent gamma-distributed random variables. Therefore,

$$\begin{aligned} p(L|\mathcal{H}_1) &= \frac{\kappa^{J-2M} L^{M-1}}{\Gamma(M)(\kappa+\lambda)^{J-M}} \sum_{m=0}^{M-1} \binom{M-1}{m} \frac{\Gamma(J) \kappa^m \lambda^m}{\Gamma(J-M) (\kappa+\lambda)^m L^m} \exp\left(-\frac{L}{\kappa}\right), \quad L \geq 0, \\ &= \frac{\lambda^{-J} (-L)^{J-M-1}}{\Gamma(J-M) (\kappa+\lambda)^M} \sum_{m=0}^{J-M-1} (-1)^m \binom{J-M-1}{m} \frac{\Gamma(M+m) \kappa^m \lambda^m}{\Gamma(M) (\kappa+\lambda)^m L^m} \exp\left(\frac{L}{\lambda}\right), \quad L < 0, \quad (\text{B-20a}) \end{aligned}$$

and

$$p(L|\mathcal{H}_2) = \frac{v^{J-2M} L^{M-1}}{\Gamma(M)(v+\mu)^{J-M}} \sum_{m=0}^{M-1} \binom{M-1}{m} \frac{\Gamma(J) v^m \mu^m}{\Gamma(J-M) (v+\mu)^m L^m} \exp\left(-\frac{L}{v}\right), \quad L \geq 0,$$

$$= \frac{\mu^{2M-J} (-L)^{J-M-1}}{\Gamma(J-M) (v+\mu)^M} \sum_{m=0}^{J-M-1} (-1)^m \binom{J-M-1}{m} \frac{\Gamma(M+m) v^m \mu^m}{\Gamma(M)(v+\mu)^m L^m} \exp\left(\frac{L}{\mu}\right), \quad L < 0, \quad (\text{B-20b})$$

where  $\Gamma(\cdot)$  is the gamma function.

This yields the following expressions for classifier performance:

$$P_{11}(\eta) = \sum_{m=0}^{M-1} \frac{\kappa^{J-M} \Gamma(J-M+m)}{(\kappa+\lambda)^{J-M} \Gamma(J-M)m!} \frac{\lambda^m}{(\kappa+\lambda)^m}$$

$$- \sum_{m=0}^{M-1} \frac{\kappa^{J-M} \Gamma(J-M+m)}{(\kappa+\lambda)^{J-M} \Gamma(J-M)m!} \frac{\lambda^m}{(\kappa+\lambda)^m} \gamma(M-m, \eta/\kappa), \quad \eta \geq 0,$$

$$P_{11}(\eta) = \sum_{m=0}^{J-M-1} \frac{\kappa^{J-M} \Gamma(J-M+m)}{(\kappa+\lambda)^{J-M} \Gamma(J-M)m!} \frac{\lambda^m}{(\kappa+\lambda)^m}$$

$$+ \sum_{m=0}^{J-M-1} \frac{\lambda^M \Gamma(M+m)}{(\kappa+\lambda)^M \Gamma(M)m!} \frac{1}{(J-M-1-m)!} \frac{\kappa^{J-M}}{(\kappa+\lambda)^{J-M}} \gamma(J-M-m, \eta/\lambda), \quad \eta < 0,$$

(B-21a)

while

$$P_{12}(\eta) = \sum_{m=0}^{M-1} \frac{v^{J-M}\Gamma(J-M+m)}{(v+\mu)^{J-M}\Gamma(J-M)m!} \frac{\mu^m}{(v+\mu)^m} \\ - \sum_{m=0}^{M-1} \frac{v^{J-M}\Gamma(J-M+m)}{(v+\mu)^{J-M}\Gamma(J-M)m!} \frac{\mu^m}{(v+\mu)^m} \gamma(M-m, \eta/v), \quad \eta \geq 0,$$

and

$$P_{12}(\eta) = \sum_{m=0}^{J-M-1} \frac{v^{J-M}\Gamma(J-M+m)}{(v+\mu)^{J-M}\Gamma(J-M)m!} \frac{\mu^m}{(v+\mu)^m} \\ + \sum_{m=0}^{J-M-1} \frac{\mu^M\Gamma(M+m)}{(v+\mu)^M\Gamma(M)m!} \frac{1}{(J-M-1-m)!} \frac{v^{J-M}}{(v+\mu)^{J-M}} \gamma(J-M-\mu, \eta/\mu), \quad \eta < 0,$$

(B-21b)

where  $\gamma(a, y) = \int_0^y \exp(-t) t^{a-1} dt$  is the incomplete gamma function.

Figures B-5 through B-8 plot the performance curves for representative values of  $\beta$ ,  $\psi$ ,  $\gamma$ ,  $M$ , and  $J$ . In figure B-5, all parameters except  $\gamma$  (reverberation) are held constant. In this example,  $\gamma$  runs from a minimum value of zero to its limiting value of  $\psi$ . As expected, classifier performance degrades with increased reverberation.

In figure B-6, reverberation is set to zero ( $\gamma = 0$ ),  $M$  and  $\beta$  are held constant, and  $J$  is increased (thus  $\psi$  decreases). As in case 1 (figure B-2), the diversity phenomenon is evident. For a given  $P_{12}$ , there is a particular set of targets taken from the two classes characterized by constant (and now equal) receive signal-to-noise ratios ( $M\beta$  or  $J\psi$ ) for which the spreading of this received energy over some  $J'$  channels provides for best performance.

In figure B-7, reverberation is set equal to a constant value ( $\gamma = \psi/4$ ),  $M$  and  $\beta$  are held constant, and  $J$  is increased (these parameters are comparable to those in figure B-6). When

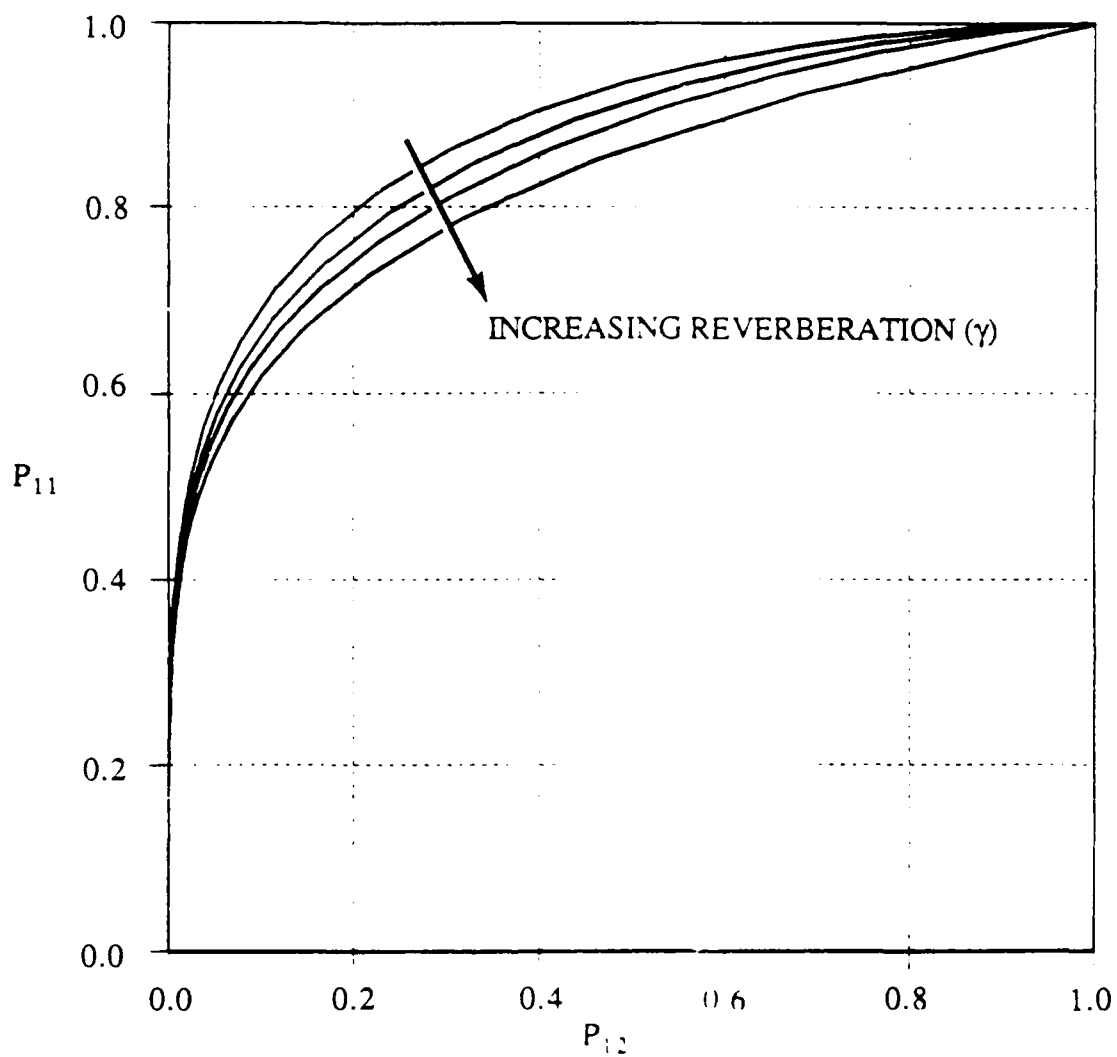


Figure B-5. Receiver Operating Characteristic (ROC) Curve for Case 2:  
 $M = 1$ ;  $J = 20$ ;  $\beta = 5$ ;  $\Psi = 0.25$ ; vary  $\gamma = 0, 0.05, 0.1, 0.25$

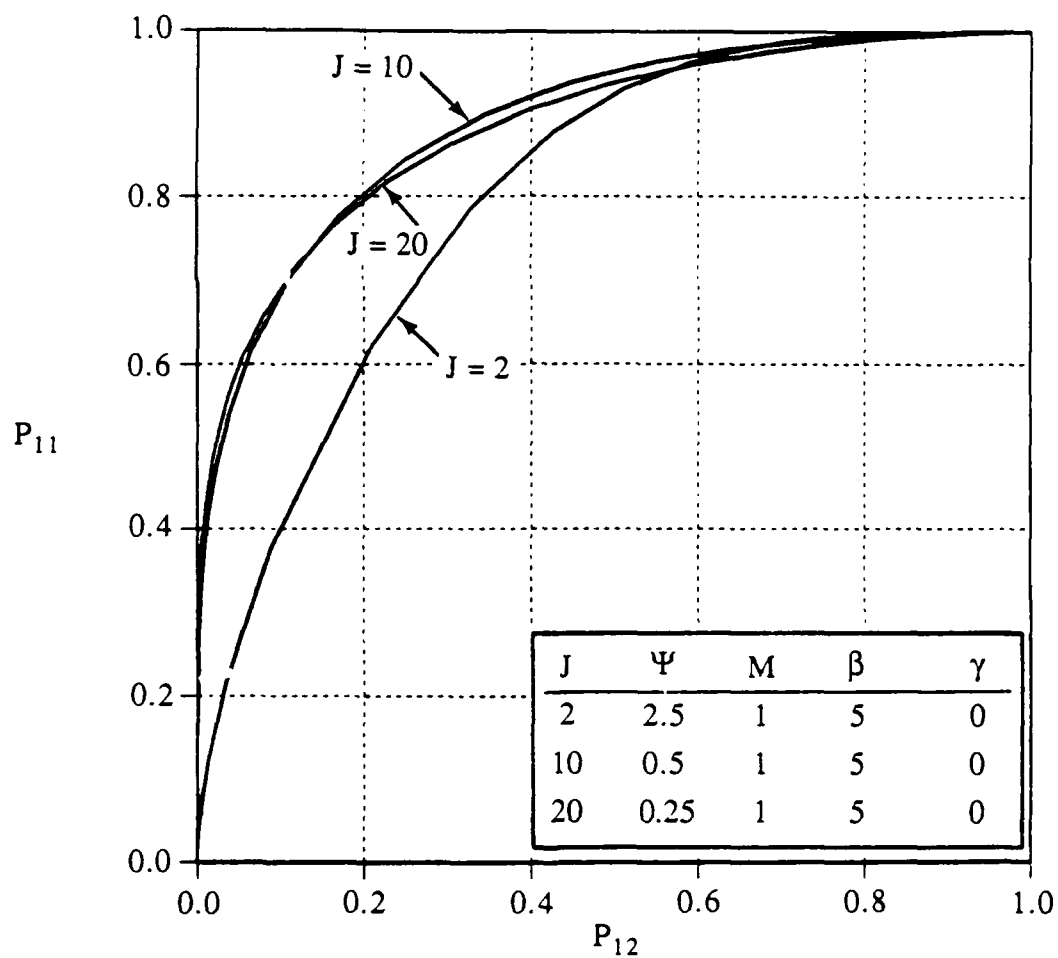


Figure B-6. Receiver Operating Characteristic (ROC) Curve for Case 2;  
 $M\beta = J\Psi = 5$ ;  $M = 1$ ;  $\gamma = 0$ , vary  $J = 2, 10, 20$



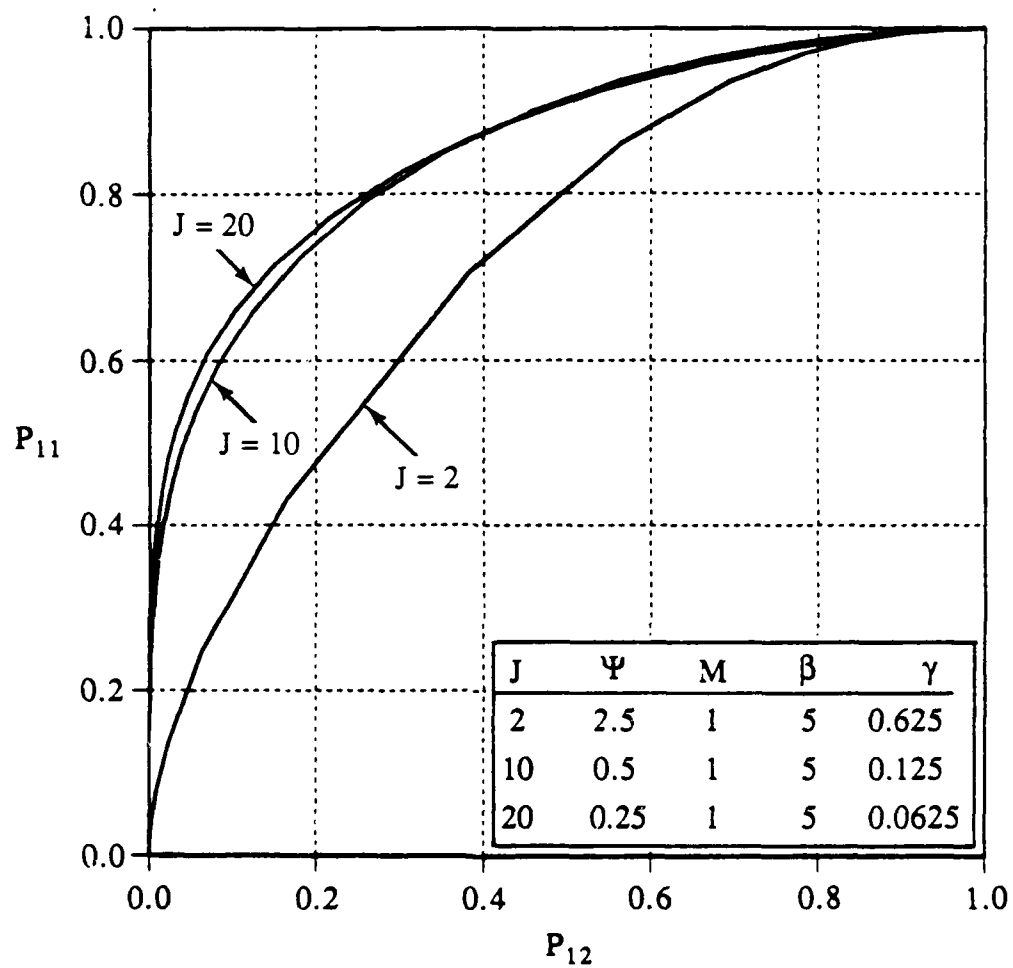


Figure B-7. Receiver Operating Characteristic (ROC) Curve for Case 2;  
 $M\beta = J\Psi = 5$ ;  $M = 1$ ;  $\gamma = \Psi/4$ ; vary  $J = 2, 10, 20$

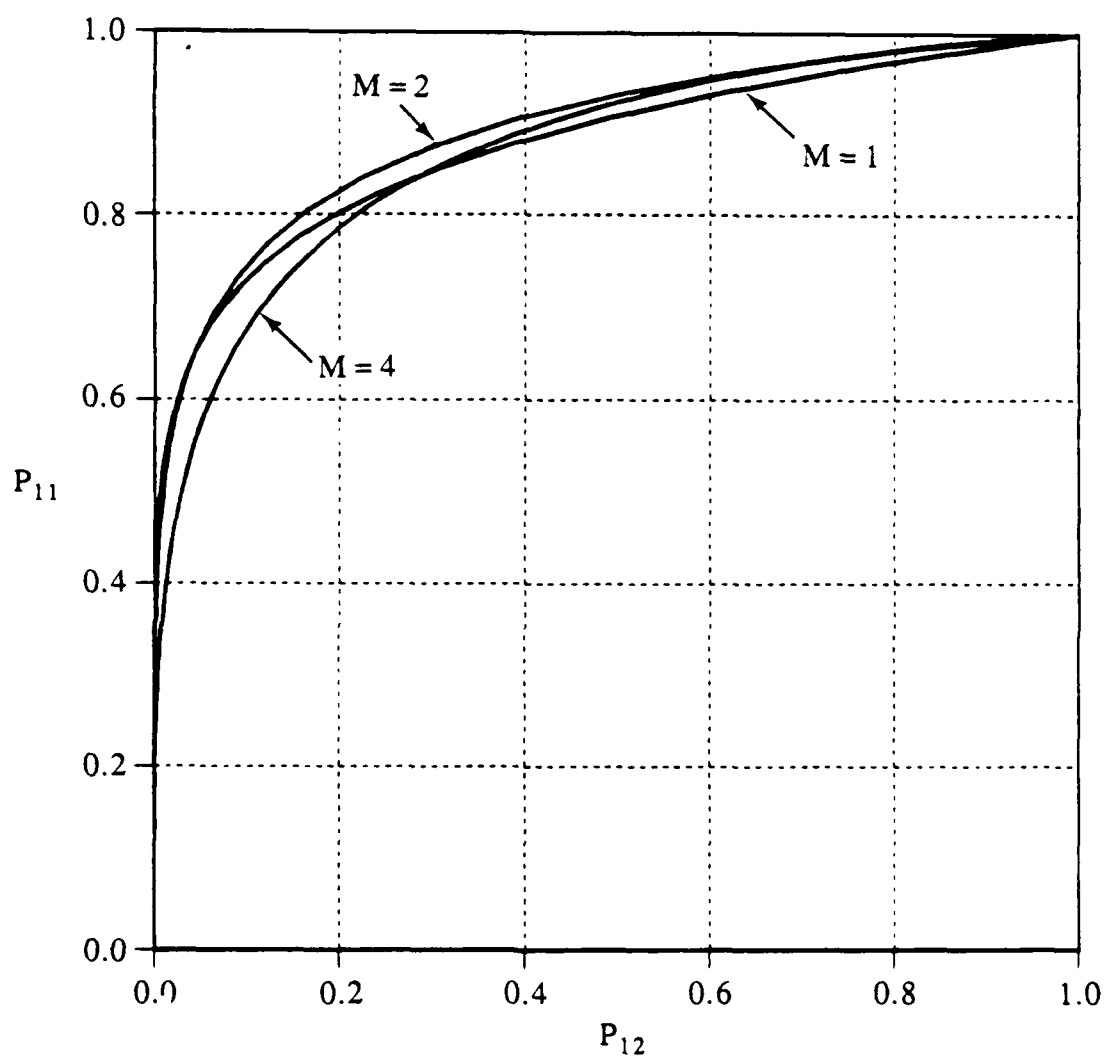


Figure B-8. Receiver Operating Characteristic (ROC) Curve for Case 2;  
 $M\beta = J\Psi = 10$ ;  $J = 20$ ;  $\gamma = \Psi$ ; vary  $M = 1, 2, 4$

compared with figure B-6, it can be seen that for a given  $P_{12}$  optimum performance can shift to a new set of targets within the class.

In figure B-8, reverberation is set to its limiting value ( $\gamma = \psi$ ),  $J$  and  $\psi$  are held constant, and  $M$  is increased (thus  $\beta$  decreases). In this example, hypothesis 2 corresponds to reverberation, while the target in hypothesis 1 spreads its fixed receive energy over an increasing number of cells. The diversity effect implies that, for a given  $P_{12}$ , there is particular target (from among the class characterized by constant receive signal-to-noise ratio) that is most easily detected in the reverberation.

## SUMMARY

Performance measures for the binary classification problem have been addressed. The sufficient statistic was shown to be a Hermitian form in the Gaussian vector  $y$ . A method for determining the necessary probability density functions via characteristic functions was given. In general, the determination of the characteristic functions requires a numerical approach; therefore, certain assumptions were made in order to present some general results. A number of examples were presented for the case of a sonar system capable of infinite resolution and targets characterized by independent scatterers. In many cases diversity was present. That is, for a given probability of misclassification, there was a best partitioning of the fixed total receive energy of a target among the cells in the test volume.

A treatment of the general problem (finite resolution, correlated scatterers) will provide further insight into the performance capability of the optimum processor. This is the subject of a related investigation, the results of which will be reported in the near future.

## REFERENCES

1. H. L. Van Trees, Detection, Estimation, and Modulation Theory, Volume 1, John Wiley and Sons Inc., New York, 1968.
2. H. L. Van Trees, Detection Estimation and Modulation Theory, Volume 3, John Wiley and Sons Inc., New York, 1968.
3. H.L. Van Trees, "A Unified Theory for Optimum Array Processing", Report 4160866, A.D.Little Inc., August 1966.
4. A B. Baggeroer and C. S. Stradling, "Joint Active and Passive Sonar Signal Processing Using Arrays," Report TP121, Naval Undersea Research and Development Center (now Naval Ocean Systems Center), San Diego, CA, December 1968 (UNCLASSIFIED). [Available to authorized requestors only.]
5. E. J. Kelly and W. L. Root, "A Representation of Vector-Valued Random Processes," Journal of Physics - Applied Mathematics, vol 39, no. 10, October 1960, p. 211.
6. T. T. Kadota, "Optimum Reception of Many Gaussian Signals in Gaussian Noise," The Bell System Technical Journal, November 1965, p. 2187.
7. J. B. Thomas and J.K. Wolf, "On the Statistical Detection Problem for Multiple Signals," IRE Transactions on Information Theory, July 1962, p. 274.
8. G.L. Turin, "On Optimal Diversity Reception," IRE Transactions on Information Theory, July 1961, p. 154.
9. A. H. Nuttall and D. W. Hyde, "A Unified Approach to Optimum and Suboptimum Processing for Arrays," Report 992, Navy Underwater Sound Laboratory (now Naval Underwater Systems Center), New London, CT, April 1969 (UNCLASSIFIED). [Available to authorized requestors only].

10. H. Mermoz, "Filtrage Adapté et Utilisation Optimale d'une Antenne," NATO Advanced Study Institute Proceedings on Signal Processing, September 1964, p. 161.
11. F. Bryn, "Optimum Signal Processing of Three Dimensional Arrays Operating on Gaussian Signals and Noise," Journal of the Acoustical Society of America, vol 34, no. 3, March 1962, p. 289.
12. R. N. Mc Donough, "Application of the Maximum-Likelihood Method and the Maximum-Entropy Method to Array Processing," in Nonlinear Methods of Spectral Analysis, S. Hadykin, ed., Springer-Verlag, New York, 1983.
13. H. Cox, "Resolving Power and Sensitivity to Mismatch of Optimum Array Processors," Journal of the Acoustical Society of America, vol 54, no. 3, 1973, p. 771.
14. B. H. Bharucha, "A Posteriori Distributions and Detection Theory," Information and Control, vol 14, 1969, p. 98.
15. J. Capon, "Multidimensional Maximum Likelihood Processing of a Large Aperture Seismic Array," Proceedings of the IEEE, vol 55, no. 2, February 1967, p. 192.
16. T.O. Lewis and P.L. Odell, Estimation in Linear Models, Prentice-Hall Inc., Englewood Cliffs, NJ, 1971.
17. F. C. Schweppe, Uncertain Dynamic Systems, Prentice Hall Inc., Englewood Cliffs, NJ, 1973.
18. A.P. Sage and J.L. Melsa, Estimation Theory with Applications to Communications and Control, McGraw Hill Book Co., New York, 1971.
19. A. V. Balakrishnan, Communication Theory, McGraw Hill Book Co., New York, 1968.
20. C. R. Rao, Linear Statistical Inference and Its Applications, John Wiley & Sons Inc., New York, 1965.

21. T. S. Ferguson, Mathematical Statistics: A Decision Theoretic Approach, Academic Press, New York, 1967.
22. G.H. Golub and C.F. Van Loan, Matrix Computations, The Johns Hopkins University Press, Baltimore, MD, 1983.
23. N.R. Goodman, "Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (An Introduction)," Annals Mathematical Statistics, vol 34, March 1963, p. 152.
24. B. Noble and J. W. Daniel, Applied Linear Algebra, Prentice-Hall Inc., Englewood Cliffs, NJ, 1988.
25. A. H. Nuttall, Exact Performance of General Second Order Processor for Gaussian Inputs, NUSC Technical Report 7035, Naval Underwater Systems Center, New London, CT, 15 October 1983 (UNCLASSIFIED).
26. D. Middleton, "Canonically Optimum Threshold Detection," IEEE Transactions on Information Theory, April 1966, p. 230.
27. S. A. Kassam, Signal Detection in Non-Gaussian Noise, Springer-Verlag, New York, 1988.
28. D. Middleton, Topics in Communication Theory, McGraw Hill Book Co., New York, 1960.

# INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
CNR (OCNR-1125--R. Obrochta, OCNR-11250A--M. Orr, OCNR-1142--T. McKenna, OCNR-23--A. Faulstich, OCNR-232--D. Houser)	5
DARPA (Attn: Naval Technology Office)	1
NOSC (Attn: Technical Library)	1
NRL (Attn: L. Palmer, E. Wald)	2
ARL, Pennsylvania State University (Attn: E. Liszka, F. Symons, R. Ingram, D. Ricker)	4
APL, University of Washington (Attn: C. Eggan, C. Sienkiewicz)	2
MIT (Attn: Prof. A. Oppenheim)	1
Yale University (Attn: Prof. P. Schultheiss)	1
CNA	1
DTIC	12